

सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI
2180
I.A.R.I.6
GIP NLK—H 3 I.A.R.I —10-5 55—15,000

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 182

LONDON
Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N W.1

16 June 1944

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES A VOLUME 182

No. A 988 6 September 1943

PAGE

Pilgrim Trust Lecture Organization of American scientists for the war. By K. T. Compton 1

The secondary electron emission from metals in the low primary energy region. By I. Gimpel and Sir Owen Richardson, F.R S. 17

On the equation of diffusion in a turbulent medium. By W. G. L. Sutton . . 48

Evaporation from a plane, free-liquid surface into a turbulent air stream. By F. Pasquill 75

The ultra-violet spectra and electron configuration of HgF and related halide molecules. By H. G. Howell. (Plate 1) . 95

No. A 989. 16 December 1943

The three infinite harmonic series and their sums (with topical reference to the Newton and Leibniz series for π) By F. Soddy, F R S. . . . 113

Relaxation methods applied to engineering problems VIII Plane-potential problems involving specified normal gradients By R. V. Southwell, F R S. and G. Vaisey 129

The refractive index of an ionized medium. II By Sir Charles Darwin, F R S. . 152

The electron diffraction by amorphous polymers By G D. Coumoulos. (Plate 2) 166

The production of penetrating showers By L Jánossy and G. D. Rochester . 180

The scattering and polarization of fast electrons by heavy elements. By C. B. O. Mohr 189

The flame spectrum of carbon monoxide. III The cool flame By A. G. Gaydon (Plates 3, 4) 199

The refraction and dispersion of sulphuretted hydrogen, sulphur dioxide and carbon oxysulphide By H. Huxley and H. Lowery (Plate 5) 207

No. A 990. 1 March 1944

Anniversary Address by Sir Henry Dale, P.R S. 217

Matrix theory of correlations in a lattice. Part I. By R. Eisenschitz . . 244

Matrix theory of correlations in a lattice. Part II. By R. Eisenschitz . . 260

The heat of adsorption of long-chain compounds and their effect on boundary lubrication. By J. J. Frewing 270

to be found there at every turn. A very important account of the methods of ventilating mines in Belgium comes next to the careful description of some animal monstrosity such as a calf with more than the proper allowance of heads or legs. The professor of astronomy at Oxford writes on the use of incubators in Egypt. Early discussions on tide prediction, then a very important matter because the old harbours did not easily accommodate the large boats wanted for the American trade, lie side by side with a disquisition on a remarkable set of teeth and pertinent references to the properties of sugar. The hodge-podge is readily understandable. The infant was beginning to notice and ask questions.... We can well be amused at all this, remembering that the infant has now become a very sturdy youth: and that the proceedings of these early fellows of the Royal Society were only the first results of the new regard for natural knowledge.'

Sir William then proceeded in most interesting fashion to sketch the significant facts in the life, work and social viewpoint of Count Rumford and Thomas Bernard, and pointed out that the wealth of Bernard and the stimulation given by Benjamin Thompson, both of them men who spent their early lives within a few miles of my home in Massachusetts, led to the establishment of your Royal Institution.

Then, after drawing lessons from the life of Michael Faraday, Sir William concluded with these words:

'Science was not merely a collection of inventions to be applied by the rich for the comfort of the poor. It was a glorious purpose to be shared by all mankind. We must try to understand the world in which we live, for our own enjoyment, for the training of our minds, for the enrichment of our souls' contemplation, for the means whereby we may help each other. It is in this spirit that we may try to do our work. It is true of course that we must work for ourselves, instructors have to earn their living and students must come to learn how to earn theirs As the world is made it must be so, but also the world is so made that the vision of its wonder and of the delight of mutual service and the happy task of exploring the one to help us in achieving the other, can light our lives for us as the sunshine lights the earth.'

As we think of the noble ideals which Count Rumford, Thomas Bernard and Michael Faraday held for science as that which 'can light our lives for us as the sunshine lights the earth', it is a grim and discouraging contrast to see the scientists of the world engaged to-day in developing new instrumentalities for destruction or other instrumentalities for protection against the destruction which would be wrought upon us by the engines of war of our enemies. The fact that this is so is a grim reminder that our skill in statesmanship and our art and ethics of Christian living have not kept pace with our ideals. We have no alternative now but to apply our knowledge of science in every aspect to serve us in our struggle for survival and to preserve for us that opportunity for which our race has struggled throughout the centuries—the opportunity to live and work in peace and freedom.

As the second personal note, let me lay claim to scientific kinship with your body as one who might be called a scientific grandchild of another one of your late

leaders, Sir Joseph Thomson. We in America affectionately called him 'J.J.', as I understand you also did in England, and we look upon him as the progenitor of that tribe of physicists who interested themselves in the conduction of electricity through gases. Taking him as the founder of that tribe, one of the second generation, your Sir Owen Richardson, was my guide and inspiration during my graduate student years at Princeton University For his sake I am sorry that I did not turn out to be a more productive pupil, but the interests and satisfactions I have had in the field of research I owe more to him than to any other man.

As you know, we in America have two principal scientific societies which are broadly representative of all the fields of science and which are rather parallel to two of your principal scientific bodies, the Royal Society of London and the British Association for the Advancement of Science

Dr Frank B. Jewett, President of our National Academy of Sciences, has asked me to deliver his personal message to the President of the Royal Society of London, Sir Henry Dale, and in addition he requested me to express to you the admiration which is felt by the members of the National Academy of Sciences for the magnificent manner in which the scientists of Great Britain have thrown the whole weight of their energies and abilities to master the innumerable technical problems arising in this war. He wanted me to assure you that in so far as we can do likewise, we in America are making a sincere effort to handle our similar problems and co-operatively to supplement the great work which you are doing.

Dr Isaiah Bowman, last month elected president of the American Association for the Advancement of Science, also gave me a message of greeting to British scientists from which I quote as follows

'Now that the war has advanced to the stage at which we begin to talk of post-war plans, we feel more than ever the need for collaboration between Great Britain and the United States While there is no such thing as an Anglo-American bloc in world politics there is such a thing as close comradeship in the fight for principles. This comradeship we feel whenever we deal with the leaders of Britain and whatever the field of interest I venture to predict that, whatever difficulties may arise, we shall find that comradeship and agreement upon principles will ever mark our future relations. This belief is based upon our widely recognized common responsibility for the peace and safety of mankind in the years after the war. If England is being changed by the war the United States is changing just as rapidly. Once our President was able to report on "the state of the Union", as our Constitution provides, almost without touching on foreign affairs. Two world wars have changed both the tenor and the scope of such messages. The state of the Union now includes the state of the world This conception of the state of the Union lays new obligations upon us all. The scientist can no longer report on the state of the sciences. He must report on the impact of science upon society. He must make use of the qualities of mind that science fosters in dealing rationally with the terrible waste in vital resources that war imposes upon the human species. We may hope that the day will soon come when every mature man and woman

will feel himself responsible for the state of the Union and act responsibly in that sector of our common life committed to his care, no matter how small the sector may be, no matter how humble.

'We say these things while recalling again how great an inheritance they represent from the unfolding political life of Great Britain. No one can speak of liberty and political responsibility and community enterprise without echoing truths that were discovered by centuries of experiment and experience on the part of rulers and ruled in the English political system. Thus, no matter where one starts in estimating future problems and future responsibilities, one ends by recognizing the special bond between America and Britain, by acknowledging the rich inheritance that has been responsible for so many strong elements in American life, and by elevating the comradeship that we both feel and need.'

And now I come to the subject of my address. I have chosen to speak on a subject pertinent to the war, to describe to you the manner in which American scientists have organized to make their contribution to the same cause which has mobilized your efforts. In so doing, I trust that I shall be within the spirit, if not the letter, of Sir William's directive for these lectures For he said, 'Such lectures would associate workers in a common task', and surely our common task right now is to direct our scientific resources for victory

In these days, when numbers of scientists are crossing the Atlantic in both directions on special missions, a better understanding of each other's organizations may be practically helpful. For I frankly confess sympathy with one of your number who recently told me that he found the American organization of interrelated scientific groups a bit complicated. I can only draw cold comfort from the fact that, complicated as it is, the scientific organization is far simpler than that of our governmental departments and bureaux generally But that is another story.

Peace-time organization of scientists in the U S.

Let me first give an over-all picture of the scientific and technical organizations of the United States as they exist in peace-time, After this brief review I shall pass to a discussion of the special scientific organizations for war, which is the subject of more particular interest to us at this time.

The scientific and engineering work in the United States may be discussed under three categories· first, the agencies of the Federal Government, exclusive of the Armed Services, secondly, the agencies within the Armed Services, and thirdly, the non-governmental agencies.

Federal Bureaux The scientific services of the Federal Government in peace-time are spread through about forty federal bureaux, of which eighteen can be called primarily scientific. Their operations involve only about half of 1% of the total peace-time federal budget, but their work is absolutely essential to the national welfare in agriculture, manufacture, commerce, health and safety. The personnel of all these bureaux operates under the Civil Service.

From the point of view of size of personnel and budget, the scientific services under the Department of Agriculture stand first in the list Probably these scientific establishments, however, are not as well known generally as those of some of the other departments because their research work is quite largely spread through a great number of agricultural experimental stations distributed throughout the various states of the Union and operated co-operatively between the Federal Government and the States. Most of the bureaux in Washington are primarily of an administrative character, but there are several which also conduct centralized research, as, for example, the Bureau of Chemistry and Soils and the Food and Drug Administration. Until recently the U S Weather Bureau operated under the Department of Agriculture, but a few years ago it was transferred to the Department of Commerce, largely because the requirements of air transportation had taken the lead in demanding more accurate and refined methods of weather forecasting than those which had served reasonably well in the past to provide for the needs of agriculture.

Many of you will probably recognize some of the more important of these governmental scientific bureaux, as, for example, the National Bureau of Standards under the Department of Commerce, the Geological Survey, the Bureau of Mines, the Bureau of Mineral Statistics and Economics under the Department of the Interior, and the National Institute of Health under the U.S. Public Health Service.

Of particular interest because of its unique character is the National Advisory Committee for Aeronautics which was established during the last war and which operates three great research establishments. Until recently the work of the N.A C.A was centred in the aerodynamical research programme at Langley Field, Virginia Several years ago there was added another aerodynamical research establishment named the 'Ames Laboratory' at Moffett Field, in California, and quite recently still another large research and development establishment for aircraft engines in Cleveland, Ohio. Also, under the N A C.A , there are currently some eighty research projects being carried on at universities under contract This obviates unnecessary duplication of facilities in a government laboratory and maintains a group of university scientists and engineers in close contact with the problems of aeronautical research.

The administration of this organization is also unique among our federal scientific agencies in that its controlling body is a committee which serves without salary and has been composed of men of such high character and distinction as to render it completely free from political influence. This committee is provided with representation from the most interested branches of the Army, Navy and Governmental Departments, but the chairman and the majority control reside in a body of citizen scientists appointed by the President who, in practice, has followed the recommendations of the chairman in appointments to fill vacancies. The present chairman of the N.A.C.A. is Professor J. C. Hunsaker, head of the Departments of Mechanical and Aeronautical Engineering at the Massachusetts Institute of

Technology, and incidentally, while a very young man, the designer of the first American airplane to fly the Atlantic.

Turning now to the United States Armed Services, I can best describe their research and development work as principally a co-operative effort between the services themselves and American industrial companies, with occasional participation from the research laboratories of the technological and educational institutions.

Army. Each branch of the Army contains a technical division under which operate laboratories or arsenals in which a certain amount of research and development work is carried on, but whose activities consist for the most part of testing and proving new war materials or equipment Thus the technical staffs in the various branches of the Army and Navy have the threefold duty of planning and co-ordinating an extensive programme of research and development carried on in the industrial laboratories, of organizing and conducting research programmes in their own establishments, and of carrying on the extensive operations of proving and testing which result in the acceptance of new devices and the drafting of specifications for production orders.

Among the principal Army establishments in which such work is centred, I would mention particularly those falling under the Ordnance Department, the Signal Corps, the Chemical Warfare Service, and the Army Air Forces. The Ordnance Department operates a great proving ground at Aberdeen, at which is centred most of the proving and testing of ordnance and research on ballistics for arms of all types. In addition it operates five principal arsenals. The Watertown Arsenal is concerned principally with the manufacture of mounts for large-calibre guns and is the principal centre for research and technical service in the field of metallurgy The Picatinny Arsenal is devoted to the testing of explosives and the design and operation of pilot plants as guides to the industrial producers. The Rock Island Arsenal carries on research and development in the field of oils and lubricants The Frankford Arsenal supplements the Aberdeen Proving Ground as a testing and a development centre for small arms The Tank Arsenal in Detroit is the centre for the design and testing of tanks.

In the Signal Corps the technical division is divided into three principal branches: the Ground Signal Branch, the Electronics Branch and the Aircraft Radio Branch. The research, development and testing work carried on under the Signal Corps is divided principally between the signal laboratories at Fort Monmouth, Camp Evans, Camp Coles, Eatontown, and Toms River The Signal Corps also maintains a large co-operative establishment working with the Army Air Forces at its principal centre, Wright Field.

Until quite recently the research and development work of the Chemical Warfare Service was centred in its great Edgewood Arsenal. As the threat of war came closer, however, a few years ago, and since a very large portion of the facilities at the Edgewood Arsenal are taken up by production, the Chemical Warfare Service established a subsidiary research laboratory and took over for this purpose the

newly erected Chemical Engineering Laboratory of the Massachusetts Institute of Technology.

Despite the great expansion of the Army Air Forces, this service has continued to concentrate its research, development and testing activities at its huge establishment at Wright Field in Ohio. There are of course many other centres at which extended service testing goes on, or at which new equipment is installed in aircraft, but Wright Field remains the headquarters for the research and development work of the Army Air Forces.

The co-ordination among all these various technical services of the Army is maintained by two types of agency Within each branch of the Army is a Board which has general supervision over technical matters within that branch. Examples are the Coast Artillery Board and the Army Engineers Board. When a project has been approved by one of these boards it is next discussed by the appropriate Technical Committee, composed of members of this branch of the service, and other branches which may be concerned with the project. If this Technical Committee also approves the project it goes as a recommendation to the general staff which presumably issues the appropriate directive

Mention should be made also of the Army Medical Corps, within which a significant amount of research is conducted under the general supervision of the Surgeon-General of the Army.

Navy The Naval Observatory and the Hydrographic Office, which are under the Chief of Naval Operations, have obvious functions in research and development work. The Marine Corps does some research, but naturally depends to a large extent on the Army and the various Bureaux of the Navy. All the Bureaux of the Navy Department, Ships, Ordnance, Aeronautics, Naval Personnel, Supplies and Accounts, Medicine and Surgery, Yards and Docks, do research and development work, though naturally the matériel Bureaux conduct the greatest volume

All research work of the Navy Department is tied together through the office of the Co-ordinator of Research and Development, which office also arranges co-ordination with the Army, with other government departments, and with the numerous civilian agencies which I will mention later

Under the Bureau of Ships the Naval Research Laboratory near Washington is a centre for all matters of fundamental research including radio, electronics, chemical warfare defence, etc. The David W Taylor Model Basin, also near Washington, is the primary station for research and development of ship structures, propeller and hull design. The Naval Boiler and Turbine Laboratory at Philadelphia is concerned with all matters of boiler research, testing and design including fuel, composition, quality and nature of boiler fuels, ceramics, etc., and also for research, test and development of main propulsion turbines. The U.S. Naval Engineering Experiment Station at Annapolis, Md., is assigned all problems of research, test and development of mechanical equipment in ships other than main propulsion, and it also has a well-equipped Diesel-engine laboratory. The principal metallurgical laboratory for the Bureau of Ships is also located at the Engineering

Experiment Station. In New York there is located the Materials Test Laboratory which handles all matters of research, test and development of electrical materials and equipment, acoustical equipment, optical and navigational material and equipment, plastics and allied materials. There are Rubber and Paint Laboratories at Mare Island, California, and an inspection test laboratory in Pittsburgh, Pa., where line-production methods for chemical analyses are set up which permit a capacity of about 5000 chemical analyses per week with a minimum of personnel and equipment In addition to the above each Navy Yard is equipped with an industrial laboratory to serve the purposes of the Yards. It has been found possible to place specialized problems in some of these laboratories such as the development of chain and rope in the Boston Navy Yard. The assignment and progress, as well as general administration of all research, development, and test work, is carried out by the Bureau of Ships in Washington in order most fully and effectively to co-ordinate all work and to collect, apply, and distribute the results.

Under the Bureau of Ordnance there are the Dahlgren Proving Ground, the Naval Gun Factory at Washington, whose research department includes the Naval Ordnance Laboratory, the Naval Powder Factory near Washington, and the Newport Torpedo Station. In addition there are establishments devoted to mines, counter-mines, nets and the like.

Under the Bureau of Aeronautics there is the Naval Aircraft Factory in Philadelphia, the Cedar Point Flight Testing Field near Washington, and the Aircraft Armament Laboratory and Testing Field at Hampton Roads

Research in medicine and surgery is directed by the Research Division of the Bureau of Medicine and Surgery, using many facilities but largely those of the U.S. Naval Medical Research Institute at Bethesda near Washington, and the Medical Research Laboratories at Pensacola and New London

Civilian agencies. I pass now to the non-governmental scientific organizations in the United States, most of whose members are attached to the staffs of some 600 colleges, universities and engineering schools, some 2000 industrial research laboratories, and other specialized research institutes Do not be alarmed when I begin by saying that these comprise well over one hundred nationally recognized scientific and engineering societies, exclusive of the social sciences Of these, only a few are general in scope in the sense that they cover broadly the entire field of science. Largest of these is the American Association for the Advancement of Science, a close parallel to your British Association, with a direct membership of about 24,000 and an indirect aggregate membership of about a million through the 187 associated and affiliated societies. Of a more exclusive character and without the affiliated and associated societies are the American Philosophical Society and the American Academy of Arts and Sciences.

Unique among the scientific organizations of the United States is the National Academy of Sciences In March 1863, during a crisis of our Civil War, Congress established the National Academy of Sciences, and President Lincoln signed the Act of Incorporation. This Act specified that 'the Academy shall, whenever called

upon by any department of the governments, investigate, examine, experiment and report upon any subject of science or art, the actual expense of such investigations, examinations, experiments and reports to be paid from appropriations which may be made for the purpose'. There was also the provision in the charter that, except for the actual expenses of these activities, neither the Academy nor any member of the Academy is entitled to receive any compensation whatsoever for such services. Although the membership is legally limited to 450, the actual membership in the Academy has never exceeded its present enrolment of 350.

Outside of its services in war-times, perhaps the most noteworthy public service by the Academy was its geological and engineering investigation of the slides which at one time threatened to prevent the successful consummation of the Panama Canal. However, the utilization of the Academy by the government has been rather 'spotty'. Under some administrations the Academy has been used rather extensively, and in other administrations has been more or less forgotten by the government In this respect I believe that your Royal Society has had a more consistent role of usefulness

One inevitable characteristic of this type of organization, in which membership is considered to be the highest scientific honour of the country, is that membership, like scientific recognition, is likely to come to a man after he has passed the peak of activity in his scientific career. For this reason the Academy has been able to perform an excellent function of the 'scientific elder statesmen' variety It has zealously kept itself free from all types of political influence Its ideals have been unselfish service, integrity and scientific competence. Frequently, however, probably in the great majority of cases, when a very active research programme has to be undertaken, many of the personnel best adapted for the particular job are not found within the membership of the Academy

During the last world war in Europe, but before the United States had become a participant, President Wilson by executive order requested the National Academy of Sciences to establish the National Research Council as a measure of national preparedness. This organization operated so usefully during the war that after its termination, in April 1919, the National Research Council was perpetuated by the National Academy of Sciences at the express request of President Wilson.

This National Research Council is organized into nine permanent divisions covering the various fields of scientific research and of scientific administration. These divisions are composed of appointed members and also of representatives from many of the scientific and engineering societies and branches of the government Because of this wide representation the National Research Council is a most effective agency for finding just the right persons to do any specific scientific job.

During the present war the National Academy and the National Research Council have been called upon to perform many important services, some of an advisory character and some involving the placing of contracts for research and development work in various laboratories.

Among the nearly 200 committees operating under the National Research Council, the following are typical of those concerned with the war: Aviation Medicine, War Metallurgy, Passive Protection against Bombing, War Use of Research Facilities, Tin Smelting and Reclamation, Clothing, Shock and Transfusions, Treatment of Gas Casualties, War-time Diet, and Selection and Training of Service Personnel.

War-time scientific organization

In spite of the apparently complete peace-time organization which I have just described, it has always been our experience, in the time of great emergency, that it appears advisable to establish temporary new agencies to deal particularly with the emergency. For example, I happened to be attached to one of these temporary agencies during the last war, and I mention the matter not only by way of illustration but also because it will enable me to relate an anecdote about your late distinguished colleague, then Sir Ernest Rutherford

This agency was the Research Information Service, set up jointly by our Military Intelligence, Naval Intelligence and Council of National Defence, with offices in Washington, London, Paris and Rome. The function of these offices was essentially the same as that of the Scientific Liaison Offices which have been operating so effectively between units of the British Commonwealth and the United States during the present war.

The head of the Research Information Service in London was the late Professor Bumstead whom some of you doubtless remember I was attached to the Paris office and happened to be temporarily in charge during the time when an allied conference on submarine detection was arranged in Paris under the auspices of this office.

One of the delegates from Great Britain was Sir Ernest Rutherford who had been collaborating closely with the French physicist, Paul Langevin, in the development of underwater supersonic devices. The day before the conference, when the British and American delegations came over from London, Rutherford was not present but he sent me a letter, delivered by Professor Bumstead, stating that some very recent experiments which he and his research assistant had been carrying on in the Cavendish Laboratory had apparently indicated success in disintegrating the nucleus of the hydrogen atom. 'If this is true', Rutherford wrote, 'it is a fact of far greater importance than the war.' He went on to say that he was in the midst of a second experiment to check these startling findings and that he would be delayed a couple of days pending the termination of this experiment. Then Rutherford added as a postscript: 'Tell nobody about this because I may be mistaken.' Later it developed that what Rutherford had actually done had not been to disintegrate the hydrogen nucleus, but rather to disintegrate the nuclei of nitrogen atoms. So far as I know, Rutherford's letter to me was the first written indication of success in the long, long struggle to produce by artificial means a transmutation of one chemical element into another. I wish I had kept that letter

and had turned it over to Professor Eve at the time when he was writing his interesting biography of Ernest Rutherford. But to return to our topic.

I have frequently tried to analyse the reasons for the establishment of special scientific agencies during times of crisis. The reasons I think are varied and rather fundamental. One of them is that every great crisis involves conditions so different from the normal situation that the types of organizations which can survive and operate during peace-time are not adequate to meet the emergency It may be, for example, that the emergency calls for exercise of very extensive administrative functions, such as the supervision of research projects and the disbursement of large governmental funds to a far greater extent than in peace-time. Hence a peace-time body of scientists organized primarily to exercise advisory functions may not be organized in a manner suited to prompt and efficient executive action Another reason is the impossibility of always maintaining in the administrative positions of peace-time agencies the personnel who would be most effective for handling important projects in a war emergency Men who have the proper capabilities are frequently too busy and too active in other directions to be willing to hold down positions in a peace-time organization which is relatively inactive. Consequently when the emergency comes, the only alternatives may be to change the leadership in the existing organizations, a difficult if not impossible process, or to set up new temporary agencies to deal with the emergency.

Whatever the reasons may be, this present war emergency has run true to form and has resulted in the establishment of a group of special agencies of temporary character which I shall proceed now to describe. It is these agencies which are carrying the principal burden of the scientific research and development work related to the war, in the United States.

The National Roster of Scientific and Specialized Personnel was established early in July 1940, when President Roosevelt approved a project for making available in one central office an index of all American citizens who have special scientific or professional skill. Headed by President Leonard Carmichael of Tufts College, this agency operates under the War Man-power Commission under the Office for Emergency Management of the executive office of the President As a result of information secured from questionnaires sent to all members of all scientific and professional organizations in the country, and supplemented by other information, an elaborate punch-card system has been set up in which practically every person in the country with specialized training or skill is listed with reference to his or her major professional fields and with the addition of a great deal of supplementary information regarding special interests, languages read or spoken, foreign countries travelled in, previous experience in the armed services or in industry, etc

There are altogether fifty-nine special fields listed in the roster, falling under the general categories of administration and management, agricultural and biological sciences, engineering and related fields, humanities, medical sciences and related fields, physical sciences, and social sciences. At the present time the total number

of persons in this roster is about 600,000, including, in October last, 71,511 chemists, 7297 mathematicians, 10,080 physicists or astronomers, 4559 radio engineers, 14,729 electrical engineers, only 408 professional philosophers, and the smallest entry in the list is 142 speleologists.

As an illustration of the manner of use of this roster I quote the following paragraph from a report by Dr Carmichael:

'How would you like to get an order for the names of all Americans who possess a knowledge of epidemiology and chemotherapy, who are competent in the diagnosis and control of *Endamoeba histolytica*, and other protozoan infections, have a knowledge of the Hindustani language, are skilled in the operation and use of specialized bacteriological research apparatus and who have travelled in the tropics?' To secure this information the stops of the punch-card sorting mechanism are pushed in at the appropriate places, the cards are ground through the machine, and all of those which conform to the above specifications fall out together.

The roster was originally conceived to serve governmental agencies who might request information on scientific personnel. More recently, as serious man-power shortages have developed both in industry and in education, and as the armed services have become more and more concerned over the most effective use of all scientifically trained personnel, the roster has been used to an increasing degree in connexion with placement work and to give the supply and shortage data on professional groups. Up to the middle of last month the National Roster had certified more than 140,000 names of specialists to various agencies engaged in the war programme in the United States.

Office of Scientific Research and Development (O.S.R.D.) Most important of the scientific agencies established specially to deal with problems of this war is the Office of Scientific Research and Development, whose Director is Dr Vannevar Bush, President of the Carnegie Institution of Washington. It was created by executive order of the President in June 1941, and under it operate the National Defence Research Committee, which had been established just a year earlier, and also the more recently established Committee on Medical Research. The O.S.R.D. is directed to co-ordinate, and where necessary supplement, the scientific research and development work relating to the war among civilian agencies as well as those of the government, including the Armed Services. To facilitate this co-ordination the advisory council to the Director of O.S.R.D. includes high-ranking representatives from the War and Navy Departments, the Chairmen of the National Advisory Committee for Aeronautics, the National Defence Research Committee and the Committee on Medical Research, and, by invitation, the President of the National Academy of Sciences and the Director of the newly established Office of Production Research and Development of the War Production Board.

The principal research and development activities of the O.S.R.D. are carried on under contracts with appropriate research institutions, these contracts being financed out of an annual Congressional appropriation. At the present time these contracts involve expenditures at the rate of about $100,000,000 per year, and

there are currently active about 1400 contracts with about 200 industrial laboratories and 100 educational or special research institutions About 6000 scientists and engineers of professional grade are engaged on these contracts, with the assistance of a considerably larger number of technicians of various types

To facilitate interchange of information between O.S.R.D. and our British colleagues, an O.S.R.D. Liaison Office was established with offices in Washington and London, now headed by Dr Caryl P. Haskins and Mr Bennett Archambault, respectively. These, in co-operation with the similar liaison services of Great Britain, Canada and, less extensively, Australia and South Africa, have served well to knit together our joint scientific efforts.

The National Defence Research Committee (*N.D R.C*) operates to recommend to the Director of O.S R.D. research and development contracts in the field of instrumentalities, devices and mechanisms of warfare. Under the chairmanship of President James B. Conant of Harvard University, this Committee is composed of four civilian scientists plus one representative each from the Army and Navy, and the Commissioner of Patents Feeding into it come the recommendations from nineteen divisions, most of which are subdivided into several sections. These divisions and sections are each built around a specific functional concept such as fire control or subsurface warfare or explosives. However, there are two divisions which are in the nature of 'catch-alls'. For example, the Division of Physics and the Division of Chemistry can be defined as handling everything in these respective fields which does not fall under any one of the more sharply defined divisions

In addition to the nineteen divisions of N.D.R C. there are two panels concerned respectively with applied mathematics and engineering. The difference between a division and a panel is suggested by the fact that the Fire Control Division, for example, is concerned with the development of fire control instruments, whereas the Applied Mathematics Panel is not concerned with the development of applied mathematics as such, but rather with the use of mathematics to aid in accomplishing the objectives of the various divisions. For this reason the Applied Mathematics Panel includes membership on each divisional committee in which applied mathematics is likely to be important. The Engineering Panel serves all the divisions to expedite the transition from the stage of research and development to the stage of quantity production under Army or Navy contract.

Intimate contact between N.D R C. and its divisions on the one hand, and the Armed Services on the other, is maintained at several levels by an extensive organization of Army and Navy liaison officers who have proved invaluable as channels for acquainting N.D R.C. with the needs and desires of the Armed Services for new equipment and for making arrangements for demonstrations and service tests.

Proposals for research or development projects come to N.D.R.C. from a wide variety of sources—requests or suggestions from the Army or Navy, proposals from industrial or academic research laboratories, promising inventions transmitted to N.D.R.C. from the National Inventors' Council, or in many cases

projects originating within the N.D.R.C. committees themselves. However, the N.D.R C. has complete freedom in making its decisions on the projects which it recommends to the Director of O.S.R.D. and the priority attached to these projects, and the Director of the O.S.R.D. has complete freedom in his own judgement to authorize the recommended contracts.

For reasons of security no person serves as a member of any N.D.R.C. committee unless he has been 'cleared' by the Army and Navy Intelligence Offices, after investigation. Similarly, all personnel of the contractors working on the research and development projects are 'cleared' by these intelligence offices to whatever degree is deemed advisable in virtue of the degree of secrecy attached to the project.

The Committee on Medical Research (C.M.R.), under the chairmanship of Dr A. Newton Richards of the Medical School of the University of Pennsylvania, is in every respect parallel to the National Defence Research Committee in its organization and methods of operation. It deals exclusively with problems of war medicine, such as shock, immunization or protection against types of diseases characteristic of the present theatres of war, etc. Though considerably younger and smaller than N D.R.C. in both personnel and budget, it already has a record of substantial accomplishment.

Joint Committee on New Weapons and Equipment (J.N.W.). The organizations described thus far have proven effective in organizing and administering research projects and in maintaining close relationships and exchange of information with the Armed Services and our British allies In respect to the Armed Services, however, these relationships are primarily at the research and development level and for a time lacked one very important element necessary to make the work fully effective in the war. This missing element was an intimate relationship between the research and development agencies and the Highest Command of the Army and Navy who have the responsibility of planning the military or naval operations in which newly developed weapons might be used effectively or for which new devices should be developed. In order to fill this gap the U.S. Joint Chiefs of Staff in May 1942 established the Joint Committee on New Weapons and Equipment, composed of Dr Bush, Director of O.S R.D., as Chairman, the Assistant Chief of Staff G4 of the Army (now Brigadier-General Moses) and the Chief of the Readiness Division of the Navy (now Rear-Admiral De Laney).

J.N.W. is charged by the Chiefs of Staff with correlating the research programme of army, navy and civilian agencies. It acts through subordinate bodies of which the special mission in which I am at present engaged in England is an example.

Through J.N.W. any new weapon whose potentialities appear to be unusually significant is brought directly to the attention of the High Command for their consideration in the planning of future operations. Conversely, J.N.W. offers a direct channel through which the High Command can pass down to the research scientists a request for development of any particular instrumentality which could be particularly effective in connexion with some contemplated operation. This

type of liaison between the scientists and the High Command is new in the United States. Its possibilities are still being explored and developed, but it can be said definitely that it has already demonstrated its possibilities of great value in the war. It is a move in a desirable direction in which you have gone further than we have gone.

National Inventors' Council. War is a great stimulus to invention, not only in the research laboratories of a country, but on the part of great numbers of its citizens, some of whom are technically competent and most of whom are uninformed but sincere in their desire to be helpful. Any actively operating research organization like the O.S R D. or the Naval Research Laboratory could be quickly bogged down under the deluge of ideas and inventions induced from all sources by the war. It is very important for purposes of morale that these inventors and would-be inventors be sympathetically handled It is also important that their ideas be expertly examined to make sure that really worth-while ideas are not brushed aside, even though experience has shown that perhaps only one in one hundred thousand is new and significant.

To give such sympathetic and expert consideration and to screen the interesting suggestions out of the great mass, the National Inventors' Council was established in June 1940, in close association with the U S. Patent Office in the Department of Commerce, under the chairmanship of Dr Charles F. Kettering, Vice-President in charge of research for the General Motors Corporation. All suggested inventions relating to the war from any source and submitted to any agency or person in the government are channelled through this National Inventors' Council (unless they happen to come initially to an appropriate agency which is immediately interested in pursuing the matter). They pass through the hands of an expert staff of examiners who select those inventions which appear to have merit and bring them to the attention of the appropriate agency.

Office of Production Research and Development of the War Production Board. Until recently the organized war research efforts in the United States failed to include the very important category of research aimed at the development of substitute materials in fields where shortages exist, or of improved methods of production and manufacture. It was apparently assumed that the commercial interest of the production companies would lead them automatically to take care of this situation. However, under the pressure of war-production orders, limitations of man-power and materials, and financial regulations, the normal peace-time incentives to such research and development work by companies proved inadequate to meet the needs of the situation Consequently, last September, there was established in the War Production Board an Office of Production Research and Development under the directorship of Dr Harvey N. Davis, President of the Stevens Institute of Technology. This agency is still in the process of organization to operate somewhat along the lines of the Office of Scientific Research and Development but with primary responsibility for materials and method of production rather than for devices and instrumentalities of warfare. It is regrettable

that we did not have the foresight to establish this much-needed agency at a much earlier date, but it has already begun its operations and we hope that it may be enabled to play an important role during the balance of the war.

Engineering, Science and Management War Training Programme. Though not directly concerned with scientific research, a review of the scientific war agencies in the United States would not be complete without at least a brief reference to the efforts to increase the supply of technically trained personnel to meet the increasing demand for such personnel in every field of war activity In October 1940 a special engineering training programme was organized under the U.S. Office of Education and financed by Congressional appropriation. Later, this programme was extended to include also training in science and industrial management. It operates at both the collegiate and the technical school levels, and its magnitude may be appreciated by the fact that, even in its first year of operation, it put through its specialized courses approximately ten times as many students as graduated in that year from the regularly established engineering colleges. Most but not all of the work was carried on in night schools, and the whole programme has been decidedly helpful in relieving the technical man-power shortage.

Army and Navy Technical Training Programmes. At the present time the Army and Navy are jointly establishing a very extensive programme for the training of their own younger personnel in such fields as aeronautical engineering, naval architecture, electronics, communications, automotive engineering, etc., through contractual arrangements with several hundred of the nation's colleges and universities. Under these programmes it is anticipated that approximately 250,000 selected young men in uniform will be detailed for this training at educational institutions during the coming year, the duration of such training to vary from field to field and individual to individual, in accordance with the needs of the situation and the performance of the individual. These special collegiate programmes are intended to supplement, at the higher level, the very much larger technical training programmes which the Army and Navy are conducting in their own establishments.

Conclusion

I conclude this factual, over-long, but I hope usefully informative address on a note of faith and optimism which I am sure is shared by the allied scientists on both sides of the Atlantic Each of us concerned with some phase of the war effort is aware of some very significant new applications of scientific research in the war. For most of us, this knowledge is largely restricted to the special fields in which we ourselves have been working. Of necessity, the general public knows only in a vague way about some of these things and nothing at all about most of them.

When victory has been won, and the whole story of these scientific accomplishments can be told, it will indeed be a thrillingly interesting recital. Out of it all will come, not only its important contribution to victory, but a number of exceedingly significant results of permanent peace-time value. It is already evident that

many of these war-time developments will have very useful peace-time applications, whose contributions to our standards of living and general prosperity and comfort will help to compensate for the ravages wrought by the war. Scientists will have a renewed faith in the worth-whileness of their work, and will continue their intellectual and practical endeavours with the increased power that has come from the experience of 'team-work' on war problems. The general public, and especially the governmental and industrial leaders, will have greater appreciation of the value of science and scientists, both pure and applied—and this should result in permanently increased support of scientific research in the universities, industries and governmental agencies. These, I trust, will be some of the long-term gains to which we may look forward as the result of the temporary concentration upon practical problems of survival and victory which the war has forced upon us.

With these words of optimism, I close with the hope that the next American Pilgrim Trust lecturer to address you may not feel obliged to discuss the war, but will be able to treat of some interesting aspect of the progress of science in accord with the original conception of Sir William Bragg and as a happy feature in the post-war forward march of Science.

The secondary electron emission from metals in the low primary energy region

By Irena Gimpel and Sir Owen Richardson, F.R.S.

(*Received* 28 *October* 1942)

This paper describes a new method and apparatus for measuring and analysing the secondary electron emission from non-gaseous materials. It has been devised especially for dealing with secondary emissions caused by primary electrons of very small energy The essential features include a beam of primary electrons defined and controlled by an electrostatic electron lens system, which directs it towards the centre of a sphere A of the material investigated. A is surrounded by a concentric conducting sphere which collects the secondary electrons emitted from the surface of A. With the arrangement as set up it is possible to measure the contact potential difference between essential elements of the apparatus during the course of the experiments, without moving them or doing anything which would change the nature or composition of their surfaces.

The method is applied to the case of pure gas-free copper with primary electrons having energies down to the lowest practicable with a tungsten thermionic source (about 0 35 eV at 2000° K). The distribution of energy is analysed both for the primary and the secondary electrons. It is found to be practically the same for both groups for all energies below a few volts. From this we deduce (1) that for these low-energy electrons the secondary electrons are just reflected electrons, and (2) that the coefficient of reflexion r varies very little with the energy of the electrons. Numerous direct determinations of r have been made.

It is shown that no manipulation with fields can ever reduce the mean energy of electrons from a thermionic source below $2kT$, where T is the temperature of the source and k is Boltzmann's constant. Many determinations of r have been made from a few volts down to $2kT$, and the average value of r is about 0 24. No variation of r has been established with certainty, but there are indications that it drops a little from the value at $2kT$ to a minimum at about $2kT+0{\cdot}5$ eV, then increases slightly.

Introduction

In 1908, in an investigation of the energy distribution of thermionically emitted electrons, one of us (O. W. R.) observed that more electrons were captured by a hole than by a metal surface of equal aperture. This proved that such electrons were reflected by metal surfaces. This was found to be the case when the electrons were not accelerated by an applied electric field and had only their thermal energy of emission, kT being of the order 0·06 eV. In such a case, for a brass surface, it was found that the reflexion coefficient (ratio of number of reflected to that of incident electrons) was 0·3.

Since then a great deal has been discovered about secondary electron emission, too much to be summarized here. But we can take a typical case, a beam of 200 eV electrons. When this falls on a metal surface three groups of electrons are emitted: (1) elastically reflected electrons with energy 200 eV, (2) inelastically reflected electrons with energies under 200 eV, but with the more energetic members not much under, and (3) a group of electrons of comparatively low energy. There is evidence that group (3), or a part of it, originates in the same process as group (2). As the energy of the primary electrons is reduced from 200 eV, the proportion of the secondary electrons found in group (1) steadily increases and the proportions of those found in groups (2) and (3) steadily diminish together until with primary energies of the order 10 eV group (1) appears to contain almost the whole of the secondary electron emission.

While there is now a considerable body of evidence, on the whole fairly consistent, in support of the statements just made, there is still no general agreement either as to the character of the phenomena or the magnitude of the secondary emission for primary energies in the region of about 1 eV, or less. It might appear at a first glance that the nature of the phenomena in this small energy range is a trivial matter, but that is not at all the case, as may be judged from the following considerations.

The value of the reflexion coefficient at zero primary energy can be calculated, theoretically, from the fundamental structure of metallic surfaces. That structure is at present unknown, but the value of the coefficient is something with which the right structure must be compatible The value of the reflexion coefficient at thermionic energies also enters fundamentally into the formula for the thermionic emission. If the coefficient is zero, the temperature-independent factor in the formula is a universal constant involving Planck's h, whereas if it is unity this factor is zero and there is no thermionic emission at all, a state of things which should be attained as the contact potential of a metal tends towards $-\infty$.

The experiments to be described were carried out with the object of throwing more light on the phenomena which characterize this region.

The lack of progress in this field hitherto has not been due to want of interest or effort. It is due to the fact that one meets here with serious experimental difficulties which are absent, or at any rate unimportant, at higher energies. For a direct attack on the problem it is necessary to obtain a fairly narrow, sufficiently homogeneous

and well-defined beam of electrons. For high energies this is easy, but at low energies it tends rapidly towards impossibility at zero energy owing to the mutual repulsion of the electrons. For the same reason, fundamentally, the currents all become very small, which impairs accuracy. The contact potential differences between conducting surfaces in the apparatus become of the same order of magnitude as the volt equivalents of the impressed electron energies, and it is consequently necessary to be able to check these differences at many essential stages, since they are subject to large changes by very minute contamination.

The method of attacking the problem which we shall describe was initiated by one of us (O. W. R.) about 8 years ago and, partly owing to interruptions, the task of overcoming the various difficulties made only rather slow progress. Ursula Andrewes and M. A. El Sherbini took part in the earlier development work, and notable contributions were made later by Arthur M. Crooker. One of the authors (I. G.) has only been associated with the work since the autumn of 1938, but all the experiments and measurements actually described in the paper were made since that date, most of them in 1940 and 1941.

The work was carried out at King's College, London, until it was evacuated soon after the outbreak of war, when the apparatus was moved to the Imperial College and re-erected there. This move was probably a fortunate one, as it would almost certainly have been completely destroyed by a heavy high-explosive bomb which exploded about 50 ft away from its original site in October 1940.

The primary electron beam

A narrow beam of electrons of high energy sufficiently definite for experimental purposes may be obtained by geometrical limitations, for example, by making them pass through apertures in a series of diaphragms in field free space. With electrons of low velocity this method is inadequate

If a beam of electrons of circular cross-section is travelling along the axis of x in a field free space, and if when a certain section of it crosses the plane $x = 0$ all the electrons in it are moving with the same resultant velocity v parallel to the x axis, and if at that instant the radius of this section of it is r_0, it is not difficult to show that, owing to the mutual repulsion of the electrons, after this section has travelled a further distance x its radius will have become

$$r = r_0 \exp\left(\frac{2\pi e i x^2}{mv^3}\right)^{\frac{1}{2}}, \tag{1}$$

where e and m are the electronic charge and mass and i is the current density in the beam. In this equation e and i are in e.s.u. If we change over to i = mA/sq. cm. and V = energy in eV, it becomes

$$r = r_0 \exp 6{\cdot}88 i^{\frac{1}{2}} x / V^{\frac{1}{2}}. \tag{2}$$

Thus the lower the speed of the electrons in the beam the more it spreads out.

The path of the electrons between the cathode and the target in the apparatus used in this investigation was about 6·5 cm. long. If there were no electric field the spreading of the electron beam would be considerable in this distance, which it is not practicable to reduce very much.

This difficulty has been overcome by placing a series of six electrically charged diaphragms, each with a small circular aperture, in the path of the electrons from the source to the target. This constitutes a set of electrostatic electron lenses. A central section of this part of the apparatus, actual size, as it was used in the experiments, in shown in figure 1. It is mounted in an envelope of transparent fused silica drawn to scale in figure 2. Only a part of this, easily recognizable from the figures, is included in figure 1.

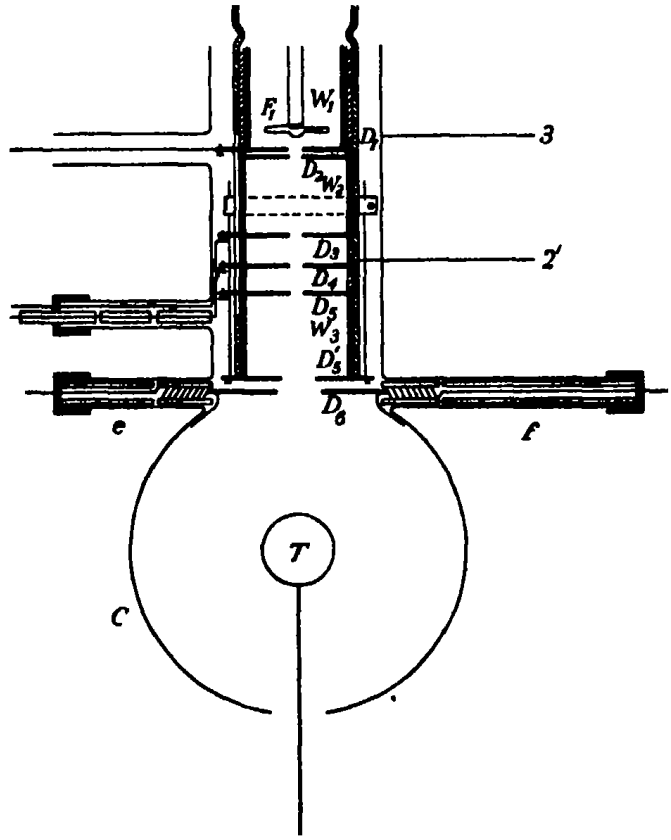


FIGURE 1

The source of primary electrons is the hot cathode F_1 (figure 1). This consisted of a tungsten strip 1·2 cm. long, 0·05 cm. wide and 0·004 cm. thick. The strip is fixed to two stout copper wires which are led outside the apparatus through two quartz tubes passing through the inner quartz stopper (1, figure 2) to which they are also fused. These quartz tubes serve as support for the leads and filament. The copper leads are crossed immediately behind the filament to minimize the magnetic field from the current. The heating current was 5 amp.

The strip was bent so as to form a V-shaped bulge in the middle of it. The bend in the V is appreciably hotter than the rest of the filament, so that owing to the very rapid increase of electron emission with rising temperature it becomes a good

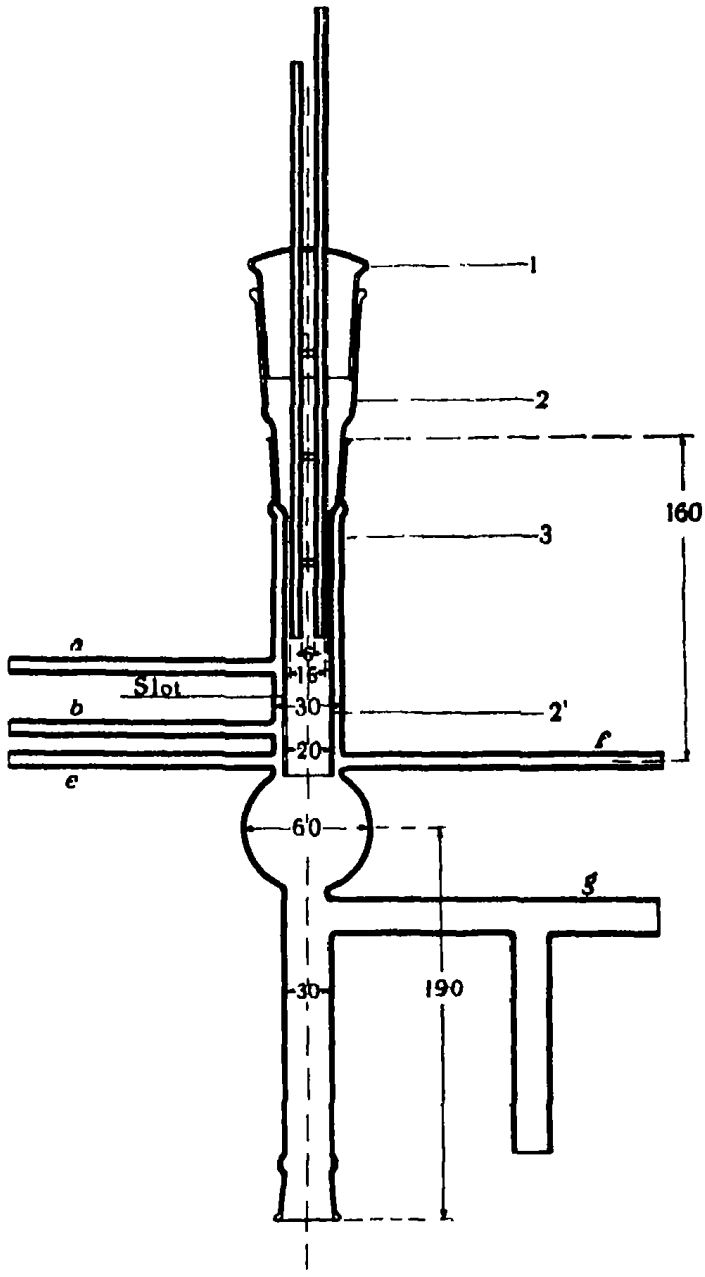


FIGURE 2

approximation to an equipotential electron source. The source also maintains a more constant position relative to the diaphragms than when the bend is absent.

The electrons from the filament F_1 (figure 1) pass through the seven circular apertures D_1, D_2, D_3, D_4, D_5, D_5' and D_6 in successive parallel copper plates and fall

on the copper target T where they are absorbed or reflected or emit secondary electrons. The secondary electrons and the reflected electrons are received by the collector C. This is a conducting sphere formed by evaporating pure copper on to the inner surface of the spherical quartz envelope. In figure 1 the target T is represented as a sphere; in the actual experiments it was part, approximately hemispherical, of a spherical sheet of pure copper and is marked T in figure 3. The common centre of the spheres T and C, the centres of the seven apertures and of the effective portion of the filament, all lie on the common axis of the cylindrical portion of the quartz envelope 3 of figure 2 (the outermost cylinder of figure 1) and of the cylindrical extension 2′ of the outer stopper 2 of figure 2 (the outermost cylinder within the envelope in figure 1) All the diaphragms except D_6 are mounted on the inside of 2′ in a manner which can be seen from figure 1.

Starting from the top of figure 1 we have a quartz tube which fits into 2′. This tube is kept in place by a constriction in the wall of the cylinder 2′. Pressing against the ground end of this quartz tube is the diaphragm D_1 which forms the plane flanged end of the copper cylindrical box W_1. W_1 slides into the quartz tube. The diaphragms D_1 and D_2, and the diaphragms D_3, D_4 and D_5, are separated from each other by quartz rings fitting into the cylinder 2′. The diaphragms D_2 and D_3 are the plane ends of the cylindrical copper box W_2 which slides into the cylinder 2′. The diaphragms D_5 and D_5' are the plane ends of the cylindrical box W_3. All the diaphragms, boxes and quartz spacers are kept in place by thin rods driven into the flange of D_5' whose other ends are attached to a copper strip which is fixed tight on to the outside of the cylinder 2′ The last diaphragm D_6 is a flat copper plate sited so as to continue as far as possible the spherical contour of C with the copper lining of which it is in electrical connexion. It is supported by tungsten springs mounted in the quartz side tubes e and f. The outside connexions to D_1, D_3, D_4 and D_5 are brought out, first through a long slot cut in the cylinder 2′ of figure 2 and then through the other two side tubes of figure 1. All the diaphragms D_1–D_6 are made of pure copper sheet 0·2 mm. thick. The circular apertures in D_1–D_5 are all 1·8 mm. in diameter, that of D_5' is 6 mm., that of D_6 is 8 mm. in diameter

As an instrument of research it is desirable that the electron lens system should be compact, robust, easily taken apart and reassembled when changes are required and also put back exactly as it was when changes are not wanted. The apparatus described has been found to fulfil these requirements satisfactorily.

Let us now consider the working of this electron lens system.

It consists of two lenses. The diaphragms D_1, D_2 form an immersion lens. The cylinder W_2 and the diaphragms D_3, D_4, D_5, the cylinder W_3 and the diaphragm D_5' make up a single lens.

It follows from the known behaviour of space-charge limited currents between two parallel plates that large thermionic currents even under low voltages can be obtained by placing the diaphragm D_1 close enough to the cathode F_1. The distance between F_1 and D_1 was varied in different experiments between 0·5 and 1·0 mm. Distances much smaller than this are not practicable.

One single diaphragm in front of the cathode acts like a divergent lens. The formula for the focal length f of such a lens (Davisson & Calbick 1932) is

$$f = 4V/(E_2 - E_1), \tag{3}$$

where V is the potential of the diaphragm with respect to the cathode, and E_1 and E_2 are the electric forces on the incident and emergent sides respectively. In the case of a single diaphragm E_2 is zero and the focal length is negative. The electron beam emerging through the diaphragm gets scattered. The addition of the diaphragm D_2 at a higher potential with respect to the cathode than D_1 prevents the electron beam from scattering.

If V_1 is the potential of D_1 and $V_1 + V_2$ is the potential of D_2, $V_1 + V_2$ is called the potential of the electron gun. The electrons after emerging from the immersion lens (D_1, D_2) pass the single lens [(W_2, D_3), D_4, (D_5, W_3, D_5')] of figure 1. W_2, D_3, D_5, W_3, D_5' are connected together and kept at the same potential as D_2. The diaphragm D_4 is kept at a higher potential with respect to the cathode than D_2. The potential V_4 of D_4 with respect to D_2 is called the focusing potential of the single lens. A characteristic property of a single lens is that the electrons enter and leave it with the same velocity.

The potential V_T of the target with respect to the cathode determines the energy of the electrons in the primary beam striking the target. When the potential V_T of the target is between 0 and 3 V, it is necessary to keep the potential of the electron gun $V_1 + V_2$ higher than V_T by a few volts. When the potential of the electron gun is equal to or less than the potential of the target V_T, the currents to the galvanometers G_1, G_2 (figure 4) are too small to be measured accurately.

This means that the electrons are first accelerated to the diaphragm D_5' by the potential $V_1 + V_2$ and then retarded between D_5' and the target by the potential difference $V_1 + V_2 - V_T$. An electron beam travelling against a retarding potential difference gets dispersed. It is therefore of advantage to have the distance between the diaphragm D_6, which completes the collector, and the preceding diaphragm D_5' short. In the apparatus used this distance was reduced to 2 mm.

The electrons are prevented from scattering between the last diaphragm D_5' of the electron gun and the target T by the collecting action of the single lens. The electrons leave the single lens as a convergent beam, so that the later dispersing action of the retarding field is counteracted The electrons missing the target, and hitting the collector directly, vitiate the results, they cause the measured secondary emission coefficient to be bigger than the true one. It is therefore essential to make sure that no electrons miss the target

The focal length f of the single lens is given by the following equation (Bouvers 1935):

$$f = d(p + p^2), \tag{4}$$

where d is the distance between the diaphragms D_3 and D_4, as well as D_4 and D_5, and $p = (V_1 + V_2)/V_4$. When the potential of the electron gun $V_1 + V_2$ is kept fixed, the focal length of the single lens can be varied by varying the focusing potential V_4.

The optimum values of V_4 are found by measuring the ratio of the current to the collector i_c to the primary current i_p as a function of V_4.

As the focusing potential V_4 is increased from zero, the measured values of the ratio i_c/i_p steadily diminish at first until a certain value of V_4 is reached, after which they begin to rise again. Thus the graphs of the measured $i_c/i_p : V_4$ have a minimum at a certain V_4. At low values of V_4 the electron beam is divergent, and some of it passes outside the target and falls directly on the collector As V_4 increases, the focal length shortens and the focusing action of the single lens brings more of the beam on to the target As V_4 is increased further the focal length may become too short, so that the beam is focused on a point in front of the target which again finds itself in a divergent beam. Ultimately, the beam will become so divergent that some of it again escapes the target and falls directly on the collector. This causes the rise in the ratio i_c/i_p after passing the minimum.

These minima are often very flat, and for practical purposes the $i_c/i_p : V_4$ curves drop down to a straight line parallel to the V_4 axis. This means that once the whole beam has been focused on to the target quite a large change in the focusing potential is necessary to move any of it off again. This is probably an indication of good alinement of the centres of the filament, of the apertures and of the target.

We have not investigated the properties of this electron lens system in very great detail, as all that is essential for our main objective is to make sure that the whole primary beam falls on the target, so that none of the electrons in it reach the collector directly, that is to say, except by reflexion from the target That condition is satisfied with the focusing potential V_4 which we apply if both when V_4 is increased and when it is diminished the measured ratio i_c/i_p either increases or is unchanged.

However, we measured i_c/i_p as a function of V_4 for various values of V_2/V_1 and of V_1+V_2 which showed that the focusing potentials at which the minimum values of i_c/i_p appeared were dependent on V_2/V_1, but the minimum value of i_c/i_p was independent of V_2/V_1.

It is advantageous to keep V_1+V_2 as low as possible because (1) the retarding potential difference between D_5' and the target, $V_1+V_2-V_T$, is then small, and (2) the probability of producing slow secondary electrons on the edges of the diaphragms is decreased.

When experiments with very slow electrons are performed, it is necessary to reduce the local magnetic field of the laboratory in the apparatus.

The radius (r cm.) of curvature of the path of an electron (energy V eV) crossing a uniform transverse magnetic field of strength H gauss is given by

$$rH = 3{\cdot}37\sqrt{V}. \tag{5}$$

The local field was reduced by two permanent magnets 30 cm. long and 1 m. apart. They were placed symmetrically with respect to the quartz apparatus in the broadside-on position in the magnetic meridian. The magnetic field was measured by a coil and a ballistic galvanometer. It was reduced to about 0·05 gauss. That this reduction was adequate is clear from the following considerations.

In the electron gun the energy of the primary electrons is of the order of several eV. From equation (5) this radius of curvature of the path of a 4 eV electron, for instance, in a transverse field of 0·05 gauss is 135 cm., whereas the whole path of the electrons between the cathode F_1 and the last diaphragm D'_5 of the electron gun in figure 1 is 4 cm. Over a path of 4 cm. a radius of curvature of 135 cm. would represent a deviation of about 0·03 cm.

The distance between D'_5 and the centre of the target T in figure 1 is 3 cm. The radius of the target is 0·6 cm. The smallest radius of curvature, r, which an outer primary electron in the space between D'_5 and T may have without missing the target is given by

$$(r+0{\cdot}1)^2+3^2=(0{\cdot}6+r)^2. \tag{6}$$

Thus $r = 8{\cdot}65$ cm. From (5) the energy of an electron with $r = 8{\cdot}65$ in a transverse field of $H = 0{\cdot}05$ gauss is 0·017. Thus some of the outer electrons which have smaller energies than 0·017 eV and a smaller radius of curvature than 8·65 cm. miss the target. Under the experimental conditions described in this paper the proportion of primary electrons with energies lower than 0·017 eV is very small, so that their presence in the primary beam cannot appreciably vitiate the measurements.

The target

This has to meet a number of requirements. Perhaps the most exacting is the necessity for heating it in vacuo to a temperature high enough to get all the gas out of it Many additional difficulties and complications are met with unless this can be done in situ.

After various trials the structure shown in figure 3 was designed, constructed and tried and found to be satisfactory. The target T is hollow in the form of a cylinder ending with a hemisphere. It is made of pure copper deposited electrolytically on a lead core. The core is melted away later. The thickness of the target is about 0·2 mm. It is heated by the tungsten spiral placed inside it. The target slides on the quartz tube Q The other end of Q is connected to a pyrex tube P by a ground-glass joint. The pyrex tube P is sealed into the quartz container (figure 2) at the conical opening at the bottom of the figure, so that the centre of the hemisphere coincides with the centre of the spherical collector.

The thermocouple $T.C.$ (figure 3) was spot welded to the target T. It consisted of a platinum wire and one of platinum + 10 % rhodium. Fine wires of 0·1 mm. diameter were used to prevent appreciable cooling at the place where the thermocouple was fixed.

The adoption of two concentric spherical surfaces for the target and collector offers many theoretical advantages. It is an arrangement which has often been used by one of us (O. W. R.) for similar purposes. It ensures that the electric field between the target and the collector shall be radial. This enables the energy of the scattered or reflected electrons to be determined directly and unambiguously. The advantages and limitations of this arrangement of the two electrodes have been discussed in

detail by Lukirsky (1924). In particular, he has shown that the resolving power, E/dE, of this system for two electron beams of energies E and $E+dE$, is of the order $(D_2/D_1)^2$, where D_2 is the diameter of the collector and D_1 that of the target. In our case $D_2 = 60$ mm and $D_1 = 12$ mm., so that E/dE is about 25 and dE/E about 0·04 or say 4 %. This resolving power could have been increased by expanding the collector or contracting the target or by both methods It is, however, believed to have been adequate for our purposes, and each of the changes would have increased the difficulty of attaining some of the other conditions which had to be satisfied.

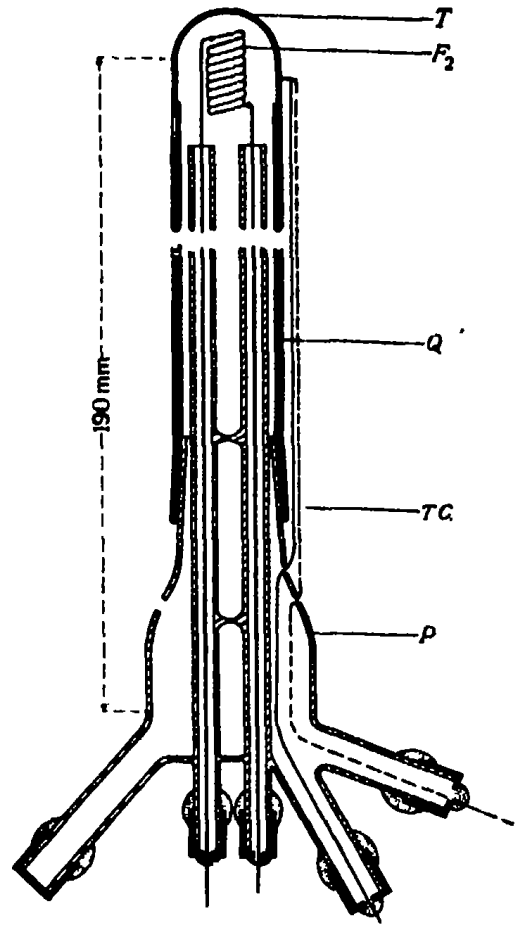


FIGURE 3

The pumping system

This was joined to the quartz container at the side tube g of figure 2. Starting from the other end it was made up of a Cenco-Hyvac pump acting as a fore-pump in series with a three-stage mercury diffusion pump of Gaede design.

The system connecting the diffusion pump with the quartz container comprises in series a liquid-air mercury trap immediately over the intake opening of the pump, a side tube leading to the MacLeod gauge, a tap with a passage of 1 in. internal diameter and two successive mercury traps. The connecting glass tubes in this high-vacuum part are all 3 cm. in diameter. The speed of the diffusion pump without connexions is 15 l /sec. at 10^{-3} mm. of mercury. The resulting speed calculated from the dimensions of the glass connexions is approximately 1 l /sec. at 10^{-3} mm. of mercury.

The trap close to the orifice catches the main part of the mercury vapour coming from the diffusion pump. The other two traps prevent traces of mercury and grease vapours from entering the quartz envelope

The MacLeod gauge, which was a sensitive one, was only used as an indicator. Readings were taken only when the pressure was so low that the mercury stuck to the top of the capillary of the gauge.

The collector

The spherical quartz shell (C_1, figure 1) was made conducting by evaporating a copper film on its inner side. The copper film was formed by the evaporation of a copper foil from a conical 15 amp. tungsten spiral.

Usually very thin films were evaporated (green in transmitted light), so that the target could be seen through them and could be placed in the centre with accuracy. A good test for the accuracy of alinement was the appearance of a very faint spot of light on the front of the copper target when the filament (F, figure 1) was lighted.

The electrical connexions

These are shown in figure 4 There is nothing very novel about the arrangements, but it is difficult to explain just what we did except with such a diagram. The ends of the filament F_1 are connected across a resistance of 3000 ohms. The centre of this resistance constitutes the zero point of the whole circuit. The current to the filament

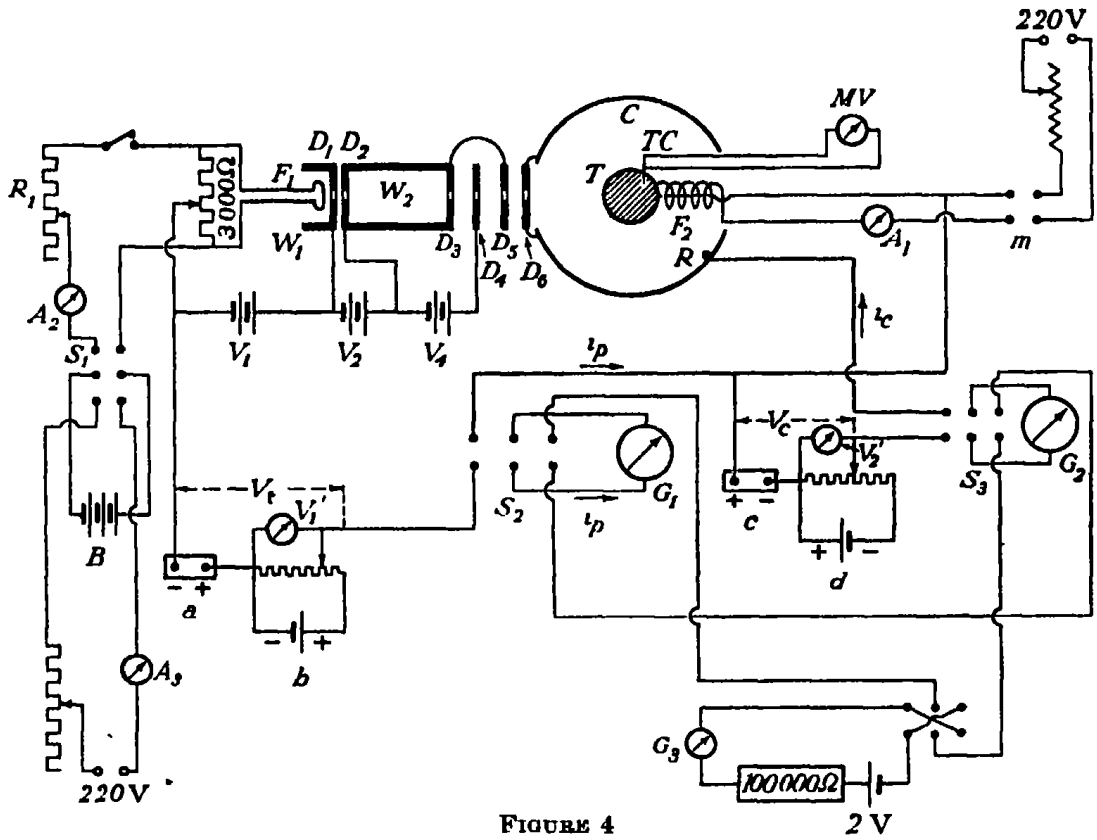


FIGURE 4

F_1 is supplied by nine cells B connected in series. The capacity of each cell is 120 amp.-hr. Cells of such high capacity were used to ensure very steady currents through the cathode filament for long periods of time The cathode heating current is controlled by the regulating resistance R_1 and measured by the ammeter A_2. The switch S_1 allows the cells to be disconnected from the cathode and connected to a charging circuit. The cells are carefully insulated.

The diaphragms D_1, (D_2, D_3, D_5) and D_4 are connected to dry batteries V_1, V_2 and V_4 respectively. The potentials of the diaphragms with respect to the cathode can be varied in 1·5 V steps.

The secondary emission coefficient is obtained from the readings of the galvanometers G_1 and G_2 which measure respectively the primary electron current i_p (as the sum of the current i_t to the target T and the current i_c to the collector C) and the secondary electron current i_c. Our arrangement measured i_c and i_p directly and simultaneously. The secondary emission coefficient is then simply the ratio i_c/i_p.

By measuring i_t as well as i_p and i_c, we could have got the same ratio as either $(i_p - i_t)/i_p$ or $i_c/(i_c + i_t)$. All three methods give the same result, but the computations take less time with the method adopted. This is an important consideration when thousands of readings have to be dealt with

The currents i_p and i_c are measured simultaneously on one scale by the two galvanometers G_1 and G_2 with sensitivities $5{\cdot}4 \times 10^{-10}$ and $3{\cdot}3 \times 10^{-10}$ amp./mm. respectively. The distance of the scale from each galvanometer is 1 m. Both are used with Ayrton Mather universal shunts. By means of the switches S_1 and S_2, G_1 and G_2 could be disconnected from the electron currents circuit and connected to a standardizing circuit which includes the galvanometer G_3. This enabled the sensitivities of both G_1 and G_2 to be checked before and after every set of readings.

The potential of the target with respect to the cathode is $V_t = (a + V_1')$ volts and that of the collector with respect to the target is $V_c = (c + V_2')$ volts, where a and c are electromotive forces of dry batteries which can be changed in steps of 1·5 V. The voltages V_1' and V_2' are obtained from two potentiometers connected to the 2 V cells b and d. The sign of V_t and V_c can be changed by interchanging the terminals of a and c respectively If, for instance, V_t has to be equal to -1 V, a is put at 1·5 V with the positive end connected to the cathode, and the potentiometer regulated so that $V_1' = +0{\cdot}5$ V.

The current to the heating circuit F_2 is supplied by the mains. When readings are taken the current in F_2 is switched off. The mains and the current controlling resistance can be disconnected from the target by means of the switch m. This switch is well insulated so that electrical leaks to earth in the circuits of the galvanometers G_1 and G_2 are avoided.

Good insulation is important generally. Wires are led separately and are passed through china insulators. The batteries V_1, V_2, V_4 and a, b, c, d are placed on a board covered with paraffin wax. That there are no leaks to earth in the whole circuit is tested by closing the circuits of the galvanometers G_1 and G_2 with all batteries connected and with no current through the filament F_1. The thermocouple $T.C.$ is connected directly to a millivoltmeter $M.V$

Experimental procedure

Usually before any readings were started the apparatus was thoroughly degassed by simultaneous heating and pumping. During this process liquid air was kept on all three traps. It was certain that no mercury penetrated into the quartz container as no sign of mercury could be detected in the trap nearest to it.

The quartz envelope and the apparatus inside it were heated by the 5 amp. tungsten ribbon F_1 and the 2 amp. spiral F_2 (figure 4). The copper target T (figures 3, 4) was degassed at temperatures above 900° C. The collector C got so hot by radiation from the target that it could not be touched by hand. In addition, the whole apparatus was heated from outside by a Bunsen burner. The first day after setting up the apparatus it was necessary to degas it for several hours before 'sticking pressure' was obtained with the target at 750° C. During the night the 1 in bore tap was closed and liquid air was kept around the middle trap. The next day only about half an hour of degassing was necessary to get sticking pressure with the target at 750° C

The routine of the measurements was as follows: The sensitivity of the galvanometers G_1, G_2 was measured, and it was checked that there were no leaks in the electrical circuit. The apparatus was then degassed for some time after sticking pressure had been reached with the target at 750° C. After this heat treatment the current in F_2 was turned off and readings were started. Readings were usually taken with the target at temperatures between 500 and 300° C. It was always found that in the range of temperature from 650 to 300° C the measured secondary emission coefficient was independent of the temperature.

The apparent coefficient was sometimes found to rise with falling temperatures of the target below 300° C. This is thought to be due to the formation of adsorbed films on the surface of the cooling target. It is for these reasons that we have relied only on measurements made between 300 and 650° C. After the target had cooled to 300° C the readings were stopped and the target was heated again to about 900° C for 10 min. After it had cooled to 650° C readings were started again. During each day the target would be heated for about 5 hr. in all.

The theory of the experiment

A. *The primary current i_p as a function of V_c for $V_c \leqslant 0$ and $V_t + E_{ft} \geqslant 0$*

The voltage i_c/i_p curves which, after differentiation, give the energy distribution of the secondaries, are obtained in the following way.

The energy eV_p of the primary electrons impinging on the target is equal to $e(V_t + E_{ft} + W_T)$, where V_t is the potential applied to the target T externally, E_{ft} is the contact potential difference between the cathode filament F and the target, and eW_T is the initial energy of emission from the cathode F. For thermionic electrons $eW_T = 2kT$, where k is Boltzmann's constant. The potential of the collector C and D_6 with respect to the cathode is $V_t + V_c + E_{fc}$ (figure 5), where V_c is the externally applied potential of the collector with respect to the target and E_{fc} is the contact potential difference between the cathode filament and the collector.

The true potential of the collector with respect to the target is

$$V_c + V_t + E_{fc} - (V_t + E_{ft}) = V_c + E_{fc} - E_{ft}.$$

When $V_c + E_{fc} - E_{fl} = 0$, all the secondary electrons emitted from the target have enough energy to reach the collector (provided the secondary emission is small enough, so that no space charge is formed round the target). When $V_c + E_{fc} - E_{fl}$ becomes negative, only those secondary electrons reach the target whose energies are greater than $e|V_c + E_{fc} - E_{fl}| = \epsilon\{|V_c| + E_{fl} - E_{fc}\}$, where $\epsilon = -e$ is now positive and $|V_c|$ is the positive numerical value of the quantity V_c.

If we plot i_c/i_p as a function of V_c with V_t constant, we obtain curves which after differentiation give the energy distribution of the secondaries.

But, as will be shown later, the $(i_c/i_p, V_c)$ curves do not correspond to a constant energy distribution of the primaries in spite of keeping V_t constant. The following discussion will show how the measured magnitudes i_c, i_p and V_c should be correlated in order to obtain curves corresponding to a constant energy distribution of the primaries.

Consider first how the potential V_0 in the opening of the diaphragm D_6 (figure 5) is a function of the potentials of the collector and of the target. It is, approximately,

$$V_0 = V_t + V_c + E_{fc} + D(V_t + E_{fl}), \tag{7}$$

where D is a positive constant depending on the geometry of the apparatus (Barkhausen 1937). V_0 is the potential which controls the magnitude of the primary current reaching the target. When V_t is kept constant, V_0 is a function of V_c only. We have to consider two cases for $V_c \leqslant 0$, namely, (1) $V_0 \geqslant 0$ and (2) $V_0 < 0$.

Case 1. The number of primary electrons arriving in front of D_6 is small, therefore i_p reaches its saturation value for $V_0 = 0$. Increasing V_0 above zero does not cause any increase of i_p. For all $V_0 \geqslant 0$ the current $i_p = i_p$ (total), i_p (total) is the saturation current,

$$V_0 \geqslant 0 \quad \text{when} \quad V_t - |V_c| + E_{fc} + D(V_t + E_{fl}) \geqslant 0,$$

or when

$$|V_c| \leqslant V_t + E_{fc} + D(V_t + E_{fl}). \tag{8}$$

For a constant V_t the energy distribution function of the primary electrons $f(V_p)$ is independent of V_c for all

$$|V_c| \leqslant V_t + E_{fc} + D(V_t + E_{fl}), \quad V_c \leqslant 0.$$

A function having a general resemblance to $f(V_p)$ is shown as a plot against V_p in figure 6. The saturation current is

$$i_p\,(\text{total}) = \int_{V_t + E_{fl}}^{\infty} f(V_p)\, dV_p, \tag{9}$$

and is represented by the area under the whole curve in figure 6.

Case 2. $V_0 < 0$.

$$V_0 < 0 \quad \text{when} \quad V_t - |V_c| + E_{fc} + D(V_t + E_{fl}) < 0,$$

or

$$|V_c| > V_t + E_{fc} + D(V_t + E_{fl}). \tag{10}$$

As V_0 assumes negative values, all electrons whose initial energy of emission $\epsilon \cdot W_t$ is $< \epsilon|V_0|$ are stopped before the diaphragm D_6 and are unable to reach the target T

(figure 5). In this case the primary current i_p is $< i_p$ (total). The energy distribution of the primaries is represented schematically on figure 6, where the shaded area represents the electrons which per unit time are stopped before D_6 and turned back towards the cathode. The unshaded area in figure 6 represents the current i_p reaching the target.

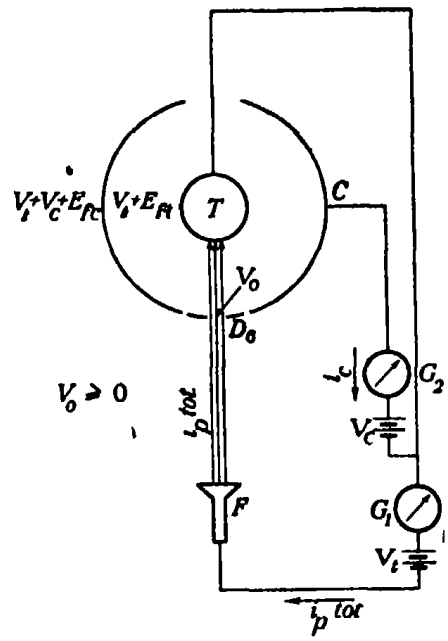


FIGURE 5

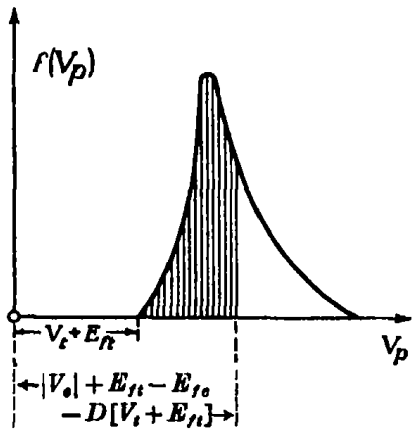


FIGURE 6

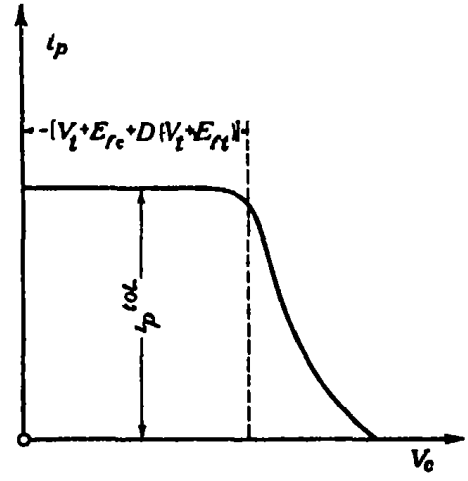


FIGURE 7

If for a given $V_i + E_{fi}$ constant and $V_0 \geq 0$ we plot i_p as a function of V_c, we get a curve like that in figure 7. The curve has a discontinuity at

$$V_c = -[V_i + E_{fc} + D(V_i + E_{fi})],$$

when the primary current begins to fall below the saturation value. Actually these discontinuities are not sharp.

When $V_i + E_{fi} = 0$, the discontinuity occurs at

$$V_c = -(V_i + E_{fc}) = E_{fi} - E_{fc}. \tag{11}$$

Thus, if we know E_{fi}, we can get E_{fc}.

B. *The energy distribution of the secondary electrons*

Let $F(V_p V_s)\,dV_s$ represent the number of secondary electrons in the energy range eV_s to $e(V_s + dV_s)$ emitted per second under the impact of unit primary current of electrons having energy eV_p.

The number of secondary electrons emitted in unit time with energies between eV_s and $e(V_s + dV_s)$ under the impact of di_p primary electrons with energies between eV_p and $e(V_p + dV_p)$ is

$$dN_c = dN_p F(V_p V_s)\,dV_s = f(V_p)\,F(V_p V_s)\,dV_s dV_p. \tag{12}$$

$F(V_p V_s) = 0$ for $V_s > V_p$, because there are no secondary electrons with energies greater than the primary energy.

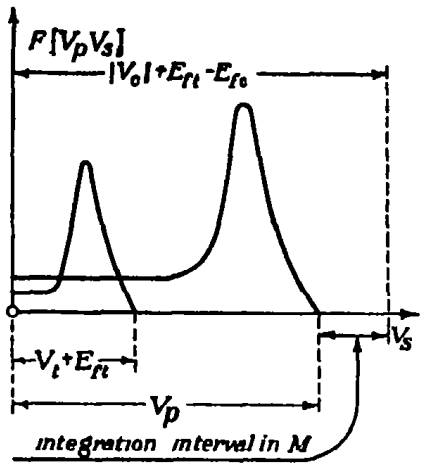


FIGURE 8

The number of secondaries having energies greater than $(|V_c| + E_{fi} - E_{fc})e$ and emitted under the impact of di_p primary electrons with energies between eV_p and $e(V_p + dV_p)$ is

$$dN_c = dN_p \int_{V_s = |V_c| + E_{fi} - E_{fc}}^{V_s = V_p} F(V_p V_s)\,dV_s. \tag{13}$$

The function $F(V_p V_s)$ for a constant V_p as a function of V_s is something like the two curves in figure 8. The value of $dN_c/dN_p = di_c/di_p$ is the area under these curves between the upper limit $V_s = V_p$ and the appropriate lower limit $V_s = |V_c| + E_{fi} - E_{fc}$. The number of secondaries having energies greater than $e(V_c + E_{fi} - E_{fc})$ emitted per second under the impact of all primary electrons hitting the target is

$$N_c = \frac{i_c}{e} = \int_{V_i + E_{fi}}^{\infty} f(V_p)\,dV_p \int_{V_s = |V_c| + E_{fi} - E_{fc}}^{V_s = V_p} F(V_p V_s)\,dV_s. \tag{14}$$

i_c is measured by the galvanometer G_2 in figure 5. This equation can be written as

$$\frac{i_c}{\epsilon} = \int_{V_i+E_{fi}}^{|V_c|+E_{fi}-E_{fc}} f(V_p)\,dV_p \int_{V_i-|V_c|+E_{fi}-E_{fc}}^{V_i-V_p} F(V_p V_s)\,dV_s + \int_{|V_c|+E_{fi}}^{\infty} \quad {}_{E_{fc}} f(V_p)\,dV_p \int_{V_i-|V_c|+E_{fi}-E_{fc}}^{V_i-V_p} F(V_p V_s)\,dV_s.$$

It can be shown that $I = \int_{V_i+E_{fi}}^{|V_c|+E_{fi}-E_{fc}} f(V_p)\,dV_p \int_{|V_c|+E_{fi}+E_{fc}}^{V_i-V_p} F(V_p V_s)\,dV_s$ is always zero.

When $|V_c| + E_{fi} - E_{fc}$ is smaller than $V_i + E_{fi}$, the integral I is zero because there are no primary electrons with energies $< V_i + E_{fi}$, so that

$$f(V_p)\,dV_p = 0 \quad \text{for} \quad V_c + E_{fi} - E_{fc} \leqslant V_p \leqslant V_i + E_{fi}$$

When $|V_c| + E_{fi} - E_{fc}$ is $> V_i + E_{fi}$, the integral I is zero because the integral

$$M = \int_{|V_c|+E_{fi}-E_{fc}}^{V_i-V_p} F(V_p V_s)\,dV_s$$

is zero for all V_p smaller than $|V_c| + E_{fi} - E_{fc}$ (see figure 8). Therefore

$$\frac{i_c}{\epsilon} = \int_{|V_c|+E_{fi}-E_{fc}}^{\infty} f(V_p)\,dV_p \int_{|V_c|+E_{fi}-E_{fc}}^{V_p} F(V_p V_s)\,dV_s. \tag{15}$$

It is seen from this equation that the current i_c consists only of secondary electrons which have been emitted under the impact of primaries with energies

$$> |V_c| + E_{fi} - E_{fc}.$$

All secondary electrons which have been emitted under the impact of primaries with energies $< |V_c| + E_{fi} - E_{fc}$ are excluded from i_c. *Therefore as far as the measurement of i_c is concerned* it is of no significance whether there are in the primary beam electrons with energies smaller than $|V_c| + E_{fi} - E_{fc}$ or not; i_c will have the same value whether the primary electron beam includes all the electrons under the whole curve in the energy distribution curve of figure 6 or only those under the unshaded portion of it.

The term $D(V_i + E_{fi})$ is small and may usually be omitted. The ratio

$$\frac{i_c}{i_p\,(\text{total})} = \int_{|V_c|+E_{fi}-E_{fc}}^{\infty} f(V_p)\,dV_p \int_{|V_c|+E_{fi}-E_{fc}}^{V_p} F(V_p V_s)\,dV_s \Big/ \int_{V_i+E_{fi}}^{\infty} f(V_p)\,dV_p$$

represents the main amount of secondary electrons with energies $> |V_c| + E_{fi} - E_{fc}$ emitted in unit time under the impact of unit primary current with an energy distribution like that represented in figure 6. If we plot i_c/i_p (total) against $|V_c|$, we obtain a curve which corresponds to a constant energy distribution of the primaries.

If, however, we plot i_c/i_p against $|V_c|$, we obtain a curve which corresponds to a constant energy distribution of the primaries for

$$0 \leqslant |V_c| \leqslant V_i + E_{fc} + D(V_i + E_{fi}),$$

because in this range of $|V_c|$ $i_p = i_p$ (total) (see figure 7). For

$$|V_c| > V_i + E_{fc} + D(V_i + E_{fi})$$

every point on the curve corresponds to a different energy distribution of the primaries.

The differentiation by $|V_c|$ of the curve got by plotting i_c/i_p (total) against $|V_c|$ leads to the main energy distribution function $A(V_s V_i)$ of the secondaries emitted under the impact of the primary current with the energy distribution as in figure 6, since, from equation (14),

$$\left.\begin{aligned}\frac{d}{d|V_c|}(i_c/i_p\,(\text{total})) \\ = \frac{d}{d|V_c|}\left(\int_{V_i+E_{fi}}^{\infty} f(V_p)\,dV_p \int_{V_s=|V_c|+E_{fi}-E_{fc}}^{V_s=V_p} F(V_p V_s)\,dV_s \Big/ \int_{V_i+E_{fi}}^{\infty} f(V_p)\,dV_p\right) \\ = \int_{V_i+E_{fi}}^{\infty} f(V_p)\,F(V_p V_s)\,dV_p \Big/ \int_{V_i+E_{fi}}^{\infty} f(V_p)\,dV_p,\ V_s = |V_c| + E_{fi} - E_{fc}.\end{aligned}\right\} \quad (16)$$

C. *The secondary emission coefficient for primary energies near zero*

If, as was the case in our experiments, E_{fc} is $> E_{fi}$, then when $V_c = 0$ the primary electrons travelling from the cathode towards the target have to pass a retarding potential difference between the diaphragm D_6 (figure 1) and the target equal to $E_{fc} - E_{fi}$. Let us consider two cases: (1) $V_i \geqslant -E_{fi}$, (2) $V_i < -E_{fi}$.

In the first case the retarding potential difference, $E_{fc} - E_{fi}$, between D_6 and the target T, is smaller than $V_i + E_{fc}$. Therefore all primary electrons arriving at D_6 can pass to the target. The current measured by the galvanometer G_1 (figure 5) is then i_p (total), the current measured by the galvanometer G_2 is

$$i_c = \epsilon \int_{V_i + E_{fi}}^{\infty} f(V_p)\,dV_p \int_{V_s=0}^{V_p} F(V_p V_s)\,dV_s$$

The ratio i_c/i_p (total) represents the true secondary emission coefficient.

In the second case the retarding potential difference, $E_{fc} - E_{fi}$, between D_6 and T is $> V_i + E_{fc}$. Therefore not all the primary electrons arriving at D_6 can pass to the target.

All primary electrons which arrive at D_6 with energies from $e(V_i + E_{fc})$ up to $e(E_{fc} - E_{fi})$ have insufficient energy to get to the target. They pass through the opening in D_6 and then are turned back to D_6 without hitting the target. All these electrons vitiate the measurement of the primary current i_p and the true secondary current i_c.

The measured secondary emission coefficient under these conditions is greater than the true one.

We see that when $V_c = 0$ and E_{fc} is $> E_{fi}$ we measure the true secondary emission coefficient only for $V_i > -E_{fi}$.

If we put $V_c = -(E_{fc} - E_{ft})$, both the collector and the target are at the same potential $V_t + E_{ft}$.

The energy distribution of the primary electrons for $V_t > -E_{ft}$ is shown on the right in figure 9. The shape of the curve is not dependent on V_t; it only gets a shift along the V_p axis if V_t changes.

For $V_t < -E_{ft}$ the shape of the curve changes as shown on the left in figure 8 The electrons which correspond to the shaded area are turned back from D_8 towards the cathode. The primary current as V_t decreases gets smaller but more uniform with respect to the electrons constituting it

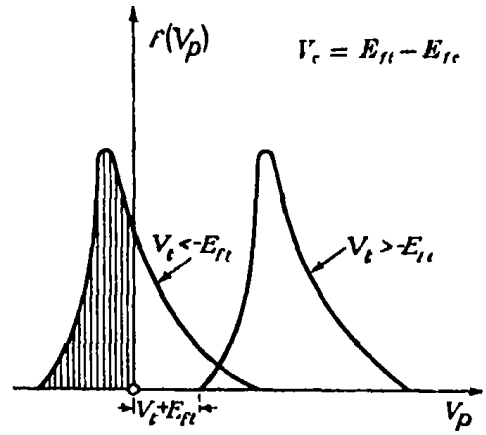


FIGURE 9

The measurement of E_{ft} and E_{fc} and of the secondary emission coefficient for primary energies near zero e-volts

The contact potential difference between the target and the cathode has been measured as follows·

The potential V_t of the target with respect to the cathode has been varied in steps from about $+1{\cdot}5$ to $-1{\cdot}5$ V, and the corresponding currents i_t have been measured. i_t has been plotted against V_t.·

The curve a of figure 10 is a typical curve obtained in this way with the hollow copper target T (figure 3) We see that the current i_t has a constant value for $V_t \geqslant 0{\cdot}3$ V. For $V_t < 0{\cdot}3$ V the current decreases rapidly with decreasing V_t.

When i_t becomes very small the electrons carrying the current are so far apart that the effect of their mutual repulsion on their motion becomes negligible; in other words, the space charge limitation of the currents disappears. In that case the rising part on the left-hand side of curves like figure 10a, b which represents the effect of the initial emission velocities and other disorganizing activities, should meet the horizontal part on the right, which represents the saturation current, at a sharp angle. This ideal is possibly attained in figure 10a, b within the experimental error, though in most of such curves a better fit with the points could be got by a slight rounding of the intersection, as, in fact, the curves are drawn in figure 10a, b. In any event, this effect, when it is as small as in these experiments, can be eliminated

by continuing the straight part of the rising curve in a straight line until it intersects the line representing the saturation current. This intersection takes place at the value of V_t for which there is no potential difference between the target and the cathode. This occurs when $V_t + E_{ft} = 0$.

The current i_t in curve *a* (figure 10) attains its saturation value for $V_t = +0{\cdot}3$ V. Therefore $E_{ft} = -0{\cdot}3$ V. The target is negative with respect to the cathode, when there is no externally applied potential difference between them.

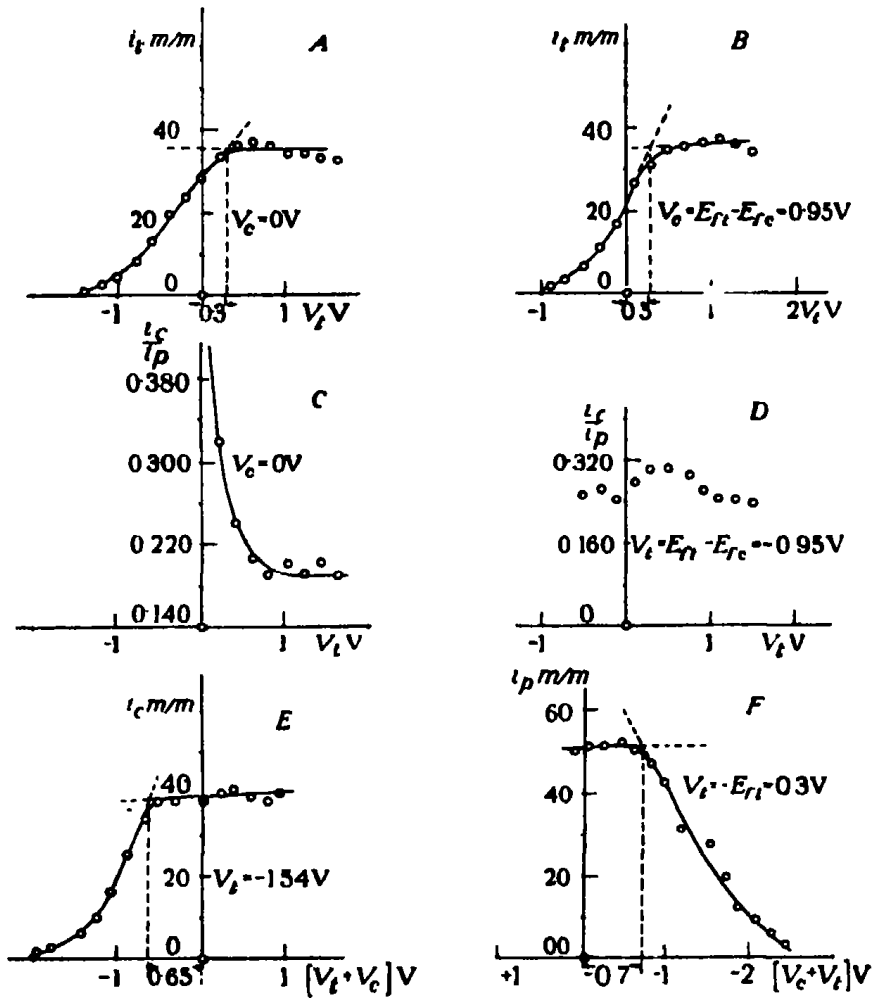


FIGURE 10

The curve *a* (figure 10) has been obtained with V_c kept at zero volts. A similar curve, *b* (figure 10), has been obtained with $V_c = E_{ft} - E_{fc}$.

The contact potential difference E_{fc} between the collector and the cathode has been measured as follows: A negative potential $V_t = -1{\cdot}54$ V was applied to the target, so that no electrons could reach it. The externally applied potential difference between the collector and the cathode $V_c + V_t = V_c - 1{\cdot}54$ was varied in steps by varying V_c and the current i_c to the collector then measured as a function of $V_c + V_t$.

We see that the break in curve *e* (figure 10) occurs at $V_c + V_t = -0{\cdot}65$ V or $V_c = +0{\cdot}89$ V. At this $V_c + V_t$ both the cathode and the collector are at the same potential, so that

$$E_{fc} = +0{\cdot}65 \text{ V}.$$

The contact potential difference between the target and the collector is

$$E_{fc} - E_{ft} = +0{\cdot}65 + 0{\cdot}3 = +0{\cdot}95 \text{ V},$$

the target being negative with respect to the collector.

The secondary emission coefficient has been measured in the range of V_t from about -1 to $+1{\cdot}5$ V. The results are shown in curves *c* and *d* of figure 10 The curve *c* has been obtained with V_c kept at zero volts and curve *d* with

$$V_c = E_{ft} - E_{fc} = -0{\cdot}95 \text{ V}.$$

It has been explained on p 34 that when $V_c = 0$ the true secondary emission coefficient is measured only for $V_t \geqslant -E_{ft}$

For $V_t < -E_{ft}$ the measured secondary emission coefficient is larger than the true one. We see from curve *c* (figure 10) that the measured secondary emission coefficient is fairly constant at about 0·2 from $V_t = +1{\cdot}6$ to $V_t = +0{\cdot}6$ V, where it begins to rise and at $V_t = -E_{ft} = +0{\cdot}3$ V it is rising rapidly as V_t falls.

Curve *d* (figure 10) represents the true secondary emission coefficient for V_t from $-0{\cdot}5$ to $+1{\cdot}5$ V because it has been obtained with $V_c = E_{ft} - E_{fc} = -0{\cdot}95$ V.

The curve *f* (figure 10), taken on a different day from the curves *a*, *b*, *c*, *d* and *e*, represents the current i_p as a function of $(V_c + V_t)$, the potential applied externally to the collector. The potential of the target with respect to the cathode, $V_t + E_{ft}$, was kept at zero It has been shown (equation (11)) that the curve i_p plotted against $V_c + V_t$ has a break at $V_c + V_t = -E_{fc}$. The break in the curve *f* occurs at $V_c + V_t = -0{\cdot}7$ V. Therefore $E_{fc} = +0{\cdot}7$ V.

The curves *e* and *f* (figure 10) thus furnish values for E_{fc} which are in good agreement

The curves shown in figure 10 are only to be regarded as samples of the observations which we have made. The number was very large and all the rest were in substantial agreement with those recorded in figure 10.

The reproducibility of the results has been very good. For example, the measurement of E_{ft} has been repeated twelve times on different days. The value $E_{ft} = -0{\cdot}3$ V was observed on nine occasions, $-0{\cdot}2$ V twice and $-0{\cdot}4$ V once

Quite similar results to those just described were obtained in a set of experiments made with the same apparatus except that the hollow target of figure 3 was replaced by a solid sphere of very pure electrolytic copper and the method of heating it was somewhat different. It was difficult to get all the gas out of this apparatus, and the results were not so closely reproducible. For example, out of eighteen determinations of E_{ft} made on separate days the value $E_{ft} = -0{\cdot}6$ V was observed eleven times, $-0{\cdot}5$ and $-0{\cdot}4$ V each twice, and $E_{ft} = 0$ three times. However, when the secondary electron coefficient i_c/i_p was determined with V_c held at the value $E_{ft} - E_{fc}$, it was found to have a constant value of about 0·16 over the range of V_t from 0 to 1 V.

The high value, about 1 V, of the contact potential difference, $E_{fc}-E_{ft}$, between the collector and the target found with the hollow target was unexpected. The contact potential difference E_{ft} between the tungsten filament and the target is known to be about right if the target has a copper surface; so we are driven to conclude that the original copper surface of the collector in the experiments with the hollow target must have got covered by something else, probably by a mixture of more volatile metals originally present in the copper of the hollow target and driven out by the heating. However, for the purpose of the experiments described here, the nature of the metal which forms the surface of the collector does not matter, so long as we are able to determine its contact potential difference with the cathode or the target.

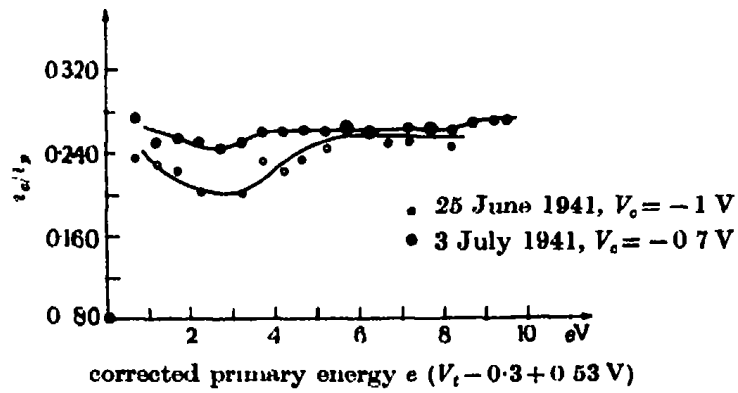


FIGURE 11.

As a matter of fact with a well-degassed copper target the secondary electron emission coefficient changes very little, not only over a range of energy of the primary electrons from 0 to 1 eV but also from 0 to at least 10 eV. Two sets of observations over this wider range taken at different dates and with different values of V_c are shown in figure 11. The abscissae give the primary energy which in these curves has been evaluated as follows:

It is clear from the curves *a*, *b*, *c* of figure 10 that the primary beam does not consist of electrons of uniform energy. There is a finite energy spread in the primary beam amounting to about 1·5 eV.

The average amount of W_d of this disorganized energy can be obtained graphically, for example, from curve *a* (figure 10) by dividing the voltage axis into 0·1 V intervals and calculating the value of

$$W_d = \frac{1}{\Delta i_t}\int_{V_t=-1}^{0} V_t\,\Delta i_t,$$

where Δi_t is the change in the current to the target in the 0·1 V interval. The value of the energy in this case is found to be about 0·72 eV.

Therefore, if the potential applied externally to the target is V_t and V_t is $> -E_{ft}$, the average energy of the electrons impinging on the target is $V_t+E_{ft}+W_d$ eV. In the curves in figure 11 W_d was 0·5 eV and $E_{ft}=-0{\cdot}3$ eV.

It will be seen that over the whole range of about 10 V covered by figure 11 the values of i_c/i_p for the upper curve only varies between the limits of 0·24 and 0·27 and for the lower only between 0·20 and 0·25. It is not certain that these fluctuations exceed the combined experimental uncertainties.

Some of the values of W_d, which varied from about 0·25 to 0·75 eV in different experiments, are higher than the mean thermal energy of the electrons emitted thermionically by the tungsten cathode which is about 0·35 eV. $2kT$ is actually 0·344 eV for a source at 2000° K. Evidently part of W_d comes from something other than this thermal energy. It is probably mainly due to the fact that, owing to the potential drop accompanying the heating current in the filament, our ideal of an equipotential electron source has only been approximately realized. The part of W_d not due to thermal energy varied, apparently irregularly, in different experiments from about $+0·1$ to about $+0·6$ eV, and, if this suggestion as to its origin is correct, these figures represent also the measure of the deviations of the source from equipotentiality. If this excess part of W_d really originates in this way, it is capable of being eliminated by appropriate modifications of the method.

Theoretically it is impossible to eliminate the energy spread of thermal origin. This remains equal to $2kT$, where k is Boltzmann's constant and T is the absolute temperature of the source, however many retarding electric filters the electrons are made to pass through. This arises from the particular nature of the velocity distribution function which ensures that at each such passage such an excess proportion of the slower electrons are removed from the electron beam by the filter as will always leave the mean energy of those which emerge equal to $2kT$ (O. W. Richardson 1909) It follows from this that in theory it is not possible to experiment with electrons from a thermionic source having less average energy than $Ve = 2kT$, which for our particular case means in practice about 0·35 eV with a tungsten cathode. With a dull emitter source this could probably be reduced to about 0·1 eV.

Let us consider what is the lowest energy for which we really have measured the secondary emission coefficient The lowest energy for which the value of the secondary emission coefficient r is given as an experimental point in curve d (figure 10) is that which corresponds to $V_f = -0\,5$ V The W_d of this is the energy under the part of curve b (figure 10) to the left of the ordinate at $V_f = -0·5$. This has been evaluated by redrawing the corresponding part of curve b (figure 10) on a larger scale and carrying out a graphical evaluation in the manner indicated on p. 38. We also have the experimental data required to determine both W_d and r for two lower energies corresponding to $V_f = -0·7$ and $V_f = -0·9$ V. The values of r for these are not plotted in curve d (figure 10).

The values of W_d and r determined for these three lowest energies are: at $V_f = -0·5$ V, $W_d = 0·31_2$ eV, $r = 0·25_2$, at $V_f = -0·7$ V, $W_d = 0·23$ eV, $r = 0·3_8$, at $V_f = -0·9$ V, $W_d = 0·14_7$ eV, $r = 0·2_9$. Naturally at these small energies no great accuracy can be claimed for the measured values as the deflexions, particularly for i_c, are very near the limit of what can be measured. But the results are very

interesting. They show that we have reached the limit of low energies which can be attained with a thermionic tungsten source.

We have seen that in a stream of thermionically emitted electrons the average kinetic energy of the electrons is $2kT$, which in practice works out for a tungsten source at about 0·35 eV. Actually at $V_t = -0{\cdot}9$ V we have found $W_d = 0{\cdot}147$ eV which is equal to half of this within the experimental uncertainty. W_d as we have evaluated it only includes half of the kinetic energy $2kT$, as it is only concerned with the kinetic energy associated with the forward component of momentum. The rest of $2kT$ comes from the two components which are associated with the two perpendicular degrees of freedom and amounts to $\frac{1}{2}kT$ each. These components are not used up in pushing the electrons against the retarding field.

This result is more general but it will be sufficient here to verify it for the case of a stream of thermionic electrons, temperature T, travelling against a uniform retarding electric field. The current density i_1 past a point where the potential is V_1 is

$$i_1 = i_0 \exp(-2hV_1 e), \tag{18}$$

where i_0 is the value of i_1 at $V_1 = 0$, $h = (2kT)^{-1}$, and e is the magnitude of the electronic charge (Richardson 1903). If i is the current at a retarding potential V numerically $> V_1$, then

$$i = i_1 \exp(-2h[V-V_1]e) = i_0 \exp(-2hVe), \quad di = -2i_0 he \exp(-2heV)\,dV. \tag{19}$$

The energy associated with the element of current di is

$$e(V-V_1)\,di = -2i_0 he^2(V-V_1)\exp(-2hVe).$$

The total of this for all the elements di which make up i_1 is

$$\int_{V=V_1}^{\infty} e(V-V_1)\,di = \int_{V_1}^{\infty} -2i_0 he^2(V-V_1)\exp(-2hVe)\,dV. \tag{20}$$

This is the equivalent of the summation (17) Putting $z = -2heV$, this becomes

$$\frac{i_0}{2h}\int_{-2heV_1}^{-\infty} -z\exp z\,dz - i_0 eV_1 \int_{-2heV_1}^{\infty} \exp z\,dz = \frac{i_0}{2h}\exp(-2heV_1) \tag{21}$$

This divided by $i_1 = (2h)^{-1} = kT$ as the value of W_d.

These experiments have furnished no evidence of any certain change in the value of the secondary electron emission coefficient r in the range from the lowest primary energy available with a tungsten source, namely, $2W_d = 2kT = 0{\cdot}35$ eV up to an energy of 1 eV or even a good deal higher. The mean of twenty-three determinations of r spread evenly over this range from about 0·3 to about 1 eV gave $r = 0{\cdot}24 \pm 0{\cdot}03$. The individual measurements ranged between 0·162 and 0·306 but only two were below 0·205 and only three above 0·276.

It is possible, of course, that there may be a discontinuous, or a very sharp, drop in r between about 0·3 eV and zero energy, but it is difficult to think of any physical reason for such an occurrence, particularly as 0·3 eV is small compared with the energy changes involved in the passage of an electron through a metallic surface.

An attempt to measure the secondary emission coefficient r at very low primary energies has been made by Farnsworth & Goerke (1930). They also used a copper target and they agreed with us in finding that after this had undergone a red-heat treatment and with no externally applied potential difference between the target and the collector, the target acquired a negative contact potential with respect to the collector. In the usual type of experiment for measuring r this leads to too high values with low energy electrons because the contact potential field deviates some of the primary electrons away from the target on to the collector, thus making the measured secondary current too high. They endeavoured to eliminate this difficulty by a procedure which they describe, and as a result of further experiments they concluded that r tends to zero as the primary energy tends to zero which is quite different from the conclusion we have reached. It seems to us that they have still not succeeded in eliminating the troublesome effects of these contact fields satisfactorily. Their method (2) as applied to the case in which both target and Faraday cylinder have been subjected to red-heat treatment requires for its success that the primary current i_{p_2} to the Faraday cylinder as measured by method (2) should be equal to the sum of the current i_{t_1} to the target as measured by method (1) and the current i_c of true secondary electrons in method (1). That $i_{p_2} = i_{t_1} + i_c$ is not at all self-evident and needs to be proved. No attempt is made to determine i_c which would involve a knowledge of the number of primary electrons which go to the collector without being reflected from the target The geometry of the electrodes in methods (1) and (2) is different; so that the electric field is different in the two cases. The primary electrons must therefore be scattered by the retarding field between the collector and the target in method (1) in a way different from that in which they are scattered by the retarding field between the collector and the target and Faraday cylinder in method (2).

These difficulties are only serious in the low-energy region. Curve 2 in figure 2 of Farnsworth & Goerke's paper, shows that at energies of 0·4 eV on their scale about half the electrons miss the target.

We shall now show that the secondary electrons generated by these primary electrons of very low energies are all elastically reflected electrons.

The energy distribution of the secondary electrons

After the hollow copper target had been degassed at temperatures above 900 and close to 1000° C (it was not safe to go higher on account of the risk of melting), the current i_c to the collector and i_p (total) were measured for various values of V_c, the potential on the collector, at a number of fixed low values of V_t, the potential on the target The ratios i_c/i_p (total) were then plotted against V_c. Typical results at $V_t = 1{\cdot}54$, 1·0, 0·7, 0·5 and 0·3 V are shown in figure 12 *a*, *b*, *c*, *d*, and *e*. The energy distribution of the secondary electrons can be obtained from these curves by graphical differentiation.

At the time when the measurements represented in the curves of figure 12 were made, the value of E_{fc}, the contact potential difference between the cathode filament and the collector, was determined by the method already described (p. 36), and found to be 0·4 V. There is no potential difference between the target and the

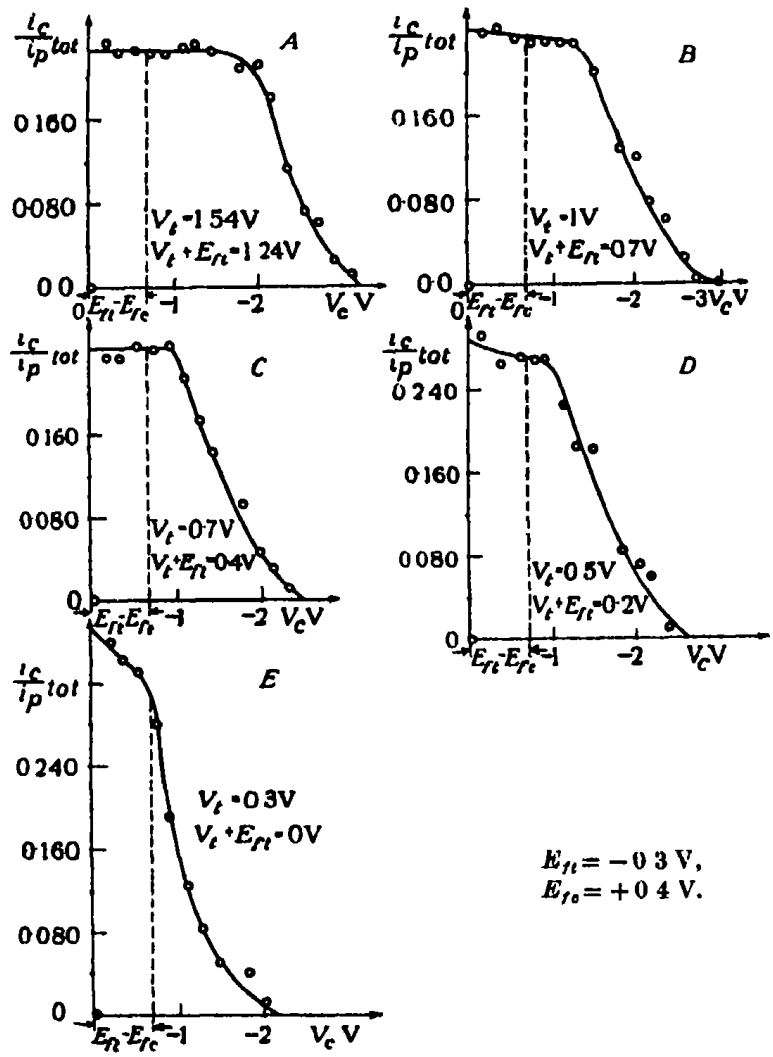


FIGURE 12

collector when $V_t + V_c + E_{fc} = V_t + E_{ft}$, where E_{ft} is the contact potential difference between the cathode and the target, or when $V_c = E_{ft} - E_{fc}$ (p. 29). Therefore the true zero when the curves of figure 12 were obtained is at

$$V_c = -0·3 - (+0·4) = -0·7 \text{ V}.$$

It is seen from these curves that all the secondary electrons have the same distribution of energy as the primary electrons. The secondary beam thus consists entirely of elastically reflected electrons.

The fact that the reflected electrons have the same energy distribution as the primaries also confirms our conclusion that the coefficient of reflexion does not vary, or at any rate varies very little, with the energy of the primary electrons at these low energies.

For $V_t = 0{\cdot}3$ V there is no potential difference between the target and the cathode, and the ratio i_c/i_p (total) decreases rapidly for $V_c < E_{ft} - E_{fc}$.

Similar results confirming these conclusions were obtained in another set of experiments during which E_{ft} was also $-0{\cdot}3$ V, but E_{fc} was $+0{\cdot}7$ V and the true zero was at $V_c = -0{\cdot}3 - (+0{\cdot}7) = -1{\cdot}0$ V.

The work function of thermally etched copper

Knowing E_{ft}, the contact potential difference between the copper target and the tungsten cathode, we can obtain the work function ϕ_{Cu} of copper from the equation

$$\phi_{Cu} = \phi_W - E_{ft}, \tag{22}$$

where ϕ_W is the work function of tungsten. Our measurements of E_{ft} have been made with the tungsten cathode at about 2000° C.

The work function of tungsten increases slowly with rising temperatures. According to Reimann (1934) it is = 4·68 V at 2000° C For the hollow copper target which had been heated to temperatures above 900° C we found $E_{ft} = -0{\cdot}3$ V Thus

$$\phi_{Cu} = 4{\cdot}68 - (-0{\cdot}3) = 4{\cdot}98 \text{ V}$$

for pure polycrystalline copper degassed at over 900° C.

It is interesting to compare our value with those got by Underwood (1935) for the {111} and {100} faces of copper single crystals. For {111} ϕ_{Cu} he found 4·86 V and for {100} ϕ_{Cu} 5·61 V The work function, 4·98 V of our target, is only slightly higher than that of the {111} face of the single crystal. It seems very likely therefore that on a polycrystalline copper surface heated at temperatures a little below its melting-point the {111} faces grow at the expense of the other ones.

This conclusion is in agreement with observations made by Dobinski (1936) He found that on a copper surface formed by solidification in vacuo or on a surface heated for a long time at 950° C there grew crystals orientated with either {111} or {100} planes parallel with the surface, the {111} orientation being the most frequent.

It is also in agreement with the theory of Kossel-Stranski (Kossel 1928; Anderson 1941) on the thermal etching of metals. According to their theory the crystal facets developed by the thermal etching of any metal of the face-centred cubic class will consist predominantly of the crystal planes of densest packing.

Comparison with theories

Since we have shown that the slow secondary electrons we are dealing with are all reflected primary electrons, the secondary emission coefficient r is in this case the same thing as the reflexion coefficient $R(W)$ of the primary electrons. Nordheim

(1928*a*,*b*) was the first to show how to calculate the reflexion coefficient of low-energy electrons at the surface of a metal on the principles of wave mechanics. The magnitude of $R(W)$ is determined by the distribution of the potential energy $U(x)$ of an electron as it passes through the metal surface, x being a co-ordinate perpendicular to the surface. The general case when $U(x)$ is any function of x is, naturally, very complicated, but it can be put in a form which shows that (1) $R(W) = 0$ when $U(x)$ is constant, (2) $R(W)$ is small when $U(x)$ varies slowly with x and increases as $U(x)$ varies more rapidly with x

There are a few cases in which fairly simple solutions have been found. One is that in which $U(x)$ changes discontinuously at the metal surface from the value zero inside the metal to a constant value C outside. C/e is called the inner potential of the metal. In this case it was shown by Nordheim (1928*a*) that

$$R(W) = \left(\frac{W^{\frac{1}{2}} - (W-C)^{\frac{1}{2}}}{W^{\frac{1}{2}} + (W-C)^{\frac{1}{2}}}\right)^2, \tag{23}$$

where

$$W = E - (p_y^2 + p_z^2)/2m, \tag{24}$$

E being the total energy of the electron and $(p_y^2 + p_z^2)/2m$ that part of the energy of the electron which comes from the components of the momentum perpendicular to the x axis, i.e. parallel to the surface. The symmetry of the equation for $R(W)$ shows that the reflexion coefficient for electrons impinging on the surface from the outside with energy $(W - C)$ is the same as that of those impinging on the surface from the inside with energy W. This symmetry of the reflexion coefficient holds also when the change in $U(x)$ is not discontinuous (Mushat & Hutchinson 1937).

It also follows from this expression for $R(W)$ that for electrons impinging from outside with zero energy $(W - C = 0)$ the reflexion coefficient is equal to 1. $R(W)$ decreases steadily from 1 to 0 as the primary energy $W - C$ of the impinging electrons increases to infinity.

In the case we are considering the graph of the potential against x is like a perpendicular cliff and is illustrated in the middle of figure 13. If we limit the discussion to examples where the top of the cliff is horizontal, the reflexion should be the greatest possible in this case. The curve in figure 13 shows the reflexion coefficient calculated for this potential distribution using the constants appropriate to copper. The height of the cliff is equal to the inner potential C which has been obtained as follows: $C = \mu + \phi$, where $\mu = \frac{h^2}{8m}\left(\frac{6n}{\pi g}\right)^{\frac{2}{3}}$, n being the number of free electrons per c.c. (assumed to be 1 per atom for copper) and g $(= 2)$ the statistical weight of an electron. The value of μ is 7·12 eV. Taking for the work function ϕ the value we obtained (p. 43) for the hollow copper target, we get $C = (7{\cdot}12 + 4{\cdot}98) = 12{\cdot}1$ eV. It is seen that $R(W - C)$ falls away from the value 1 at $W - C = 0$ very rapidly at first, and then more and more slowly but quite steadily towards the value 0 at $W - C = \infty$.

This hardly corresponds with the experimental results except that it gives a reflexion coefficient of a similar order of magnitude. This is not surprising, as it is

known that the potential field at a metallic surface cannot be as simple as this model makes it. One part of this field comes from the attraction of the electrostatic image of the electron in the metal. This has the effect of rounding off the corners at the top and bottom of the perpendicular cliff. This problem has been attacked by

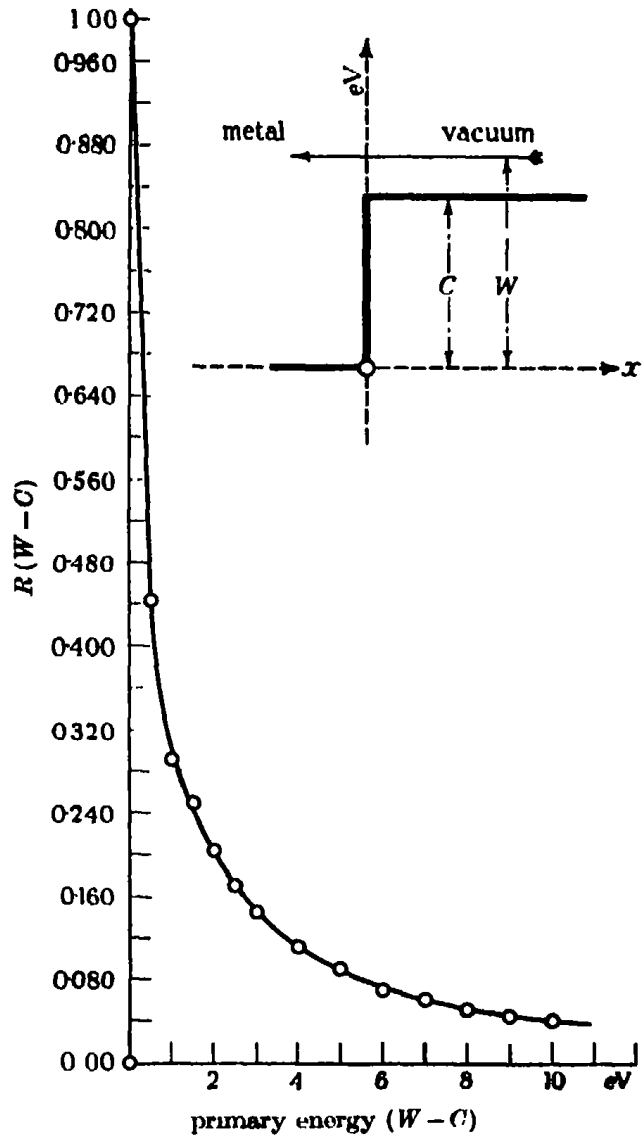


FIGURE 13. $R[W-C] = \left[\frac{W^{\frac{1}{2}}-(W-C)^{\frac{1}{2}}}{W^{\frac{1}{2}}+(W-C)^{\frac{1}{2}}}\right]^2$, $C = 12\ 1_4$V.

Nordheim (1928*b*) and by MacColl (1939), who gives numerical values. His results show that $R(W-C)$ falls away from a maximum value at $W-C=0$, as $W-C$ increases, in much the same way as for the perpendicular cliff. The chief difference is that the values of $R(W-C)$ are all much smaller (about one-tenth as large as a rule). The value of $R(W-C)$ increases with increase of the inner potential C/e.

For copper ($C = 12{\cdot}1$ eV) $R(W-C)$ is about 0·04 for $W-C=0$ and about 0·03 for $W-C=1$ eV. Our experiments have given a value of $R(W-C)$ equal to 0·24 over a range of $W-C$ from 0·2 to 1·0 eV, with no indication of a variation over this range, although a possible variation of up to 25 % or so may not be excluded with complete certainty. Furthermore, the experiments give no indication of the further diminution of $R(W-C)$ with increasing values of $W-C$ above 1 eV which is demanded by the theories. On the contrary they indicate a small rise in $R(W-C)$ in this region instead of the predicted fall.

All the theories considered assume that the potential inside the metal is a constant This assumption cannot be a valid approximation for primary energies higher than a few volts, as the electron wave-lengths then become comparable with the interatomic distances. A complete theory covering this case would have to take into account the crystalline structure of the metal and the periodic electric fields associated with it. So far as we are aware no such theory has been worked out, so we do not know what it demands of the behaviour of the reflexion coefficient. It is, however, not unlikely that the effect of the periodic fields would be to increase the reflexion coefficient at shorter wave-lengths (higher energies), and this might be enough to cancel out and reverse the monotonically falling curves given by the constant internal potential theories, as we proceed to higher energies

In describing the results we have always stated that, within the limits of experimental error, starting from $W-C=0{\cdot}2$ V, $R(W-C)$ begins by being constant and then slowly increases as $W-C$ increases After reflecting further on the matter we should like to state that what we really mean by $R(W-C)$ being constant is that it undergoes no considerable change, exceeding, let us say, about 25 % either way, in the range from $W-C=0{\cdot}2$ up to about 1 eV. We are not prepared to deny that there may be a fall in $R(W-C)$ of up to about 25 % from the value at $W-C=0{\cdot}2$ to that at $W-C=$ about 0·7 V, followed by a rise as $W-C$ increases further. If they are not regarded critically, the experiments support such a view. On the average our determinations of $R(W-C)$ do show such an initial fall followed by a rise; but, on account of the large possible experimental error, arising mainly from the very small deflexions at the lower values of $W-C$, we cannot say that they establish such a conclusion with certainty. On the other hand, they are undoubtedly compatible with it. More experiments are required to settle this issue completely.

A knowledge of the coefficient of electron reflexion in the low-energy region dealt with in this paper is important in a number of other fields which we have only room to make a list of here They include the constants of thermionic emission and allied constants of metals, for example, the energies of electron diffraction beams, inner potentials and electron refractive indices.

There are only three cases known to us where it is reasonably certain that reflexion coefficients $R(t)$ of electrons with thermionic energies have been measured The first is that of Richardson (1908, 1909) for a brass plate, not degassed The value of $R(t)$ for this was estimated to be $0{\cdot}3 \pm 0{\cdot}1$. The next was one found in some work on nickel, not yet published, by Miss Ursula Andrewes and one of us (O. W. R.).

This gave for $R(t)$ for degassed nickel $R(t) = 0{\cdot}45 \pm 0{\cdot}1$. These experiments which were made by a method different from those of either of the other two also agree with the present experiments in indicating a drop in $R(W-C)$ of about 20 % in the first 0·6 V or so from the lowest value of $W-C$ followed by a slow and steady rise up to $W-C$ = about 20 V. The value found for degassed polycrystalline copper in the present work is $R(t) = 0{\cdot}24$.

The values for $R(t)$ are not all for the same electron energy in the three cases as different thermionic sources were used. They are, however, all for values of $W-C$ = or $< 0{\cdot}4$ eV. The theories discussed above would tend to lead us to expect that nickel would have the highest $R(t)$ as its inner potential is higher than that of copper. These results, so far as they go, do not suggest any startling changes in the reflexion coefficient for different reflectors. They must, of course, all lie within the range from zero (complete absorption) to unity (total reflexion)

We had intended by suitable modifications of the apparatus to extend the experiments to a wide range of metals, but reasons connected with the war have compelled us to abandon this part of the programme, at least for the present.

In conclusion, we wish to thank Professor G. I. Finch, F.R.S., for his kindness in affording us facilities for continuing this work in his laboratory after the evacuation of King's College.

References

Anderson, P. A. 1941 *Phys. Rev* **59**, 1034.
Barkhausen, H 1937 *Lehrbuch der Elektronen-Rohren*, 4th ed. Leipzig Verlag S. Hirzel.
Bouvers, A 1935 *Physica*, **2**, 145
Davisson, C. J. & Calbick, C. J 1932 *Phys. Rev.* **42**, 580.
Dobinski, S. 1936 *Nature, Lond* , **138**, 685
Farnsworth, H. E. & Goerke, V. H 1930 *Phys. Rev.* **36**, 1191.
Kossel, W. 1928 *Leipziger Vortrage* Leipzig Hirzel.
Lukirsky, P 1924 *Z. Phys.* **22**, 351.
MacColl, L. A. 1939 *Phys. Rev.* **56**, 699.
Mushat, M. & Hutchinson, E. 1937 *Proc Nat Acad. Sci , Wash.*, **23**, 197.
Nordheim, L. W. 1928 (*a*) *Z. Phys.* **46**, 833.
Nordheim, L. W. 1928 (*b*) *Proc Roy Soc.* A, **121**, 626
Reimann, A. L. 1934 *Phys. Rev.* **45**, 898.
Richardson, O. W. 1903 *Phil. Trans.* A, **201**, 502.
Richardson, O. W. 1908 *Phil. Mag* **16**, 898.
Richardson, O. W. 1909 *Phil. Mag* **18**, 695.
Underwood, N. 1935 *Phys. Rev.* **47**, 502.

On the equation of diffusion in a turbulent medium

By W. G. L. Sutton

(*Communicated by D. Brunt, F.R.S.—Received 7 November* 1942)

The two-dimensional form of the equation of diffusion

$$u\frac{\partial\chi}{\partial x}=\frac{\partial}{\partial z}\left\{A_z\frac{\partial\chi}{\partial z}\right\} \quad (z>0),$$

under steady mean conditions in a fluid moving with mean velocity u is discussed in the case where u and A_z vary as z^m and z^{1-m} respectively $(0<m<1)$. Integrals are constructed which satisfy boundary conditions of the types arising in physical problems, and the results are applied to the theory of evaporation into a turbulent atmosphere.

1. The equation of diffusion

The two-dimensional form of the equation of diffusion, under steady mean conditions, in a fluid moving with mean velocity u is

$$u\frac{\partial\chi}{\partial x}=\frac{\partial}{\partial z}\left\{A_z\frac{\partial\chi}{\partial z}\right\}.$$

Here $\chi(z,x)$ is the value of the entity, such as vapour concentration or temperature, whose diffusion is considered, x is measured in the direction of mean flow and z in the perpendicular direction. A_z is a generalized coefficient of diffusion such that the rate per unit length at which χ is transferred across the line $z=$ *constant* in the positive direction is $-A_z\frac{\partial\chi}{\partial z}$. In non-turbulent flow A_z is the constant molecular diffusity or conductivity, and if, in addition, u is constant throughout the fluid the equation becomes that of the linear conduction of heat where x is the time co-ordinate.

In turbulent motion molecular diffusion is unimportant in comparison with that due to eddies in which a mass of fluid, small in comparison with the total mass but of dimensions large compared with the mean free path of the molecules, acts as a transporting agent. Accordingly, new forms for A_z must be sought. We consider here the form of (1·1) due to O. G. Sutton (1934), who, by a generalization of G. I. Taylor's theory of diffusion by continuous movements (Taylor 1922), obtained an expression for A_z involving a certain number n, which was regarded as specifying the degree of turbulence of the fluid, and the z-derivatives of the mean fluid velocity u. The assumption that $A_z\frac{\partial u}{\partial z}$ was constant throughout the portion of fluid considered was then shown to lead to a power law

$$u=u_1(z/z_1)^m \qquad (1\cdot2)$$

for the variation of u, where z is the distance from the (plane) boundary of the fluid and $m = n/(2-n)$. From this an explicit form

$$A_z = a u_1^{1-n} z^{1-m} \tag{1·3}$$

was obtained for A_z. In these expressions u_1 is the value of u at some standard distance z_1 and a is a certain constant involving n, z_1 and the physical constants of the fluid, but independent of u_1. It is important to observe that in the subsequent theory n, u_1 and z_1 are assumed to be independent of x, so that u and A_z are functions of z alone.

The number n is determined in any experiment from the value of m obtained by observations of the velocity profile. In the paper cited it is shown from physical considerations that $0 < n < 1$. In most technical applications, including wind-tunnel experiments, the value $n = \frac{1}{4}$ is appropriate, corresponding to the familiar velocity profile with $m = \frac{1}{7}$, but under certain meteorological conditions the value of n for the atmosphere may approach the upper bound 1.

The value of z_1 is more arbitrary. In large-scale observations, where the fluid is the atmosphere and the boundary is the earth's surface, its value is conveniently taken as 2 m. There is here no implication that z_1 represents the limit of applicability of the power law for u. But in wind-tunnel experiments this law is usually assumed to hold only within the so-called 'turbulent boundary layer' outside which the general mean velocity of the air stream is fully attained Hence if u_1 denotes this velocity z_1 must be interpreted as the breadth of the turbulent boundary layer. The condition that z_1 is independent of x implies therefore that the mathematical theory cannot strictly be applied to cases where the boundary layer is not fully developed over the whole range of values of x under consideration. This and associated questions are discussed by F. Pasquill (1943) and will be returned to later in this paper (§ 9).

If we substitute the values of u and A_z given by (1·2) and (1·3) in the differential equation (1·1) we find

$$\frac{u_1^n}{a'}\frac{\partial \chi}{\partial x} = z^{-m}\frac{\partial}{\partial z}\left(z^{1-m}\frac{\partial \chi}{\partial z}\right),$$

where $a' = a z_1^m$. In the paper cited Sutton solves this equation under boundary conditions appropriate to the case of evaporation from a saturated plane strip defined by $z = 0$, $0 < x < x_0$ It is shown by Pasquill (1943) that this solution is in good agreement with experiments on the evaporation of a wide range of liquids and also, when the physical interpretations of the symbols involved are suitably modified, with experiments on the loss of heat by forced convection from a plane surface. The equation, moreover, has proved to be a suitable basis for the discussion of evaporation problems under more general conditions than those provided by the saturated strip. It seems desirable, therefore, that a connected account of the equation should be available in a form suitable for application to such problems, and it is the aim of this paper to provide one.

For this purpose it is convenient to make a preliminary change of variables. Let x_0 be any convenient length such as that of the evaporating surface downwind. Then we write

$$x' = \frac{x}{x_0}, \quad z' = \frac{2}{2m+1}\left(\frac{u_1^n}{a'x_0}\right)^{\frac{1}{2}} z^{m+\frac{1}{2}}, \tag{1·5}$$

and obtain from (1·4) the equation*

$$\frac{\partial^2\chi}{\partial z'^2} + \frac{1}{(2m+1)z'}\frac{\partial\chi}{\partial z'} = \frac{\partial\chi}{\partial x'}.$$

If we now write $p = n/(2+n) = m/(2m+1)$,

the equation becomes

$$\frac{\partial^2\chi}{\partial z'^2} + \frac{1-2p}{z'}\frac{\partial\chi}{\partial z'} = \frac{\partial\chi}{\partial x'}, \tag{1·6}$$

which is the form in which it will be discussed.

Since $0 < n < 1$ we have $0 < p < \frac{1}{3}$, and when $n = \frac{1}{4}$ the value of p is $\frac{1}{9}$. It will be seen, however, that from a mathematical point of view the appropriate range for p is $0 < p < 1$, and this will be adopted henceforth. If $p = \frac{1}{2}$ then (1·6) becomes the familiar equation for the linear flow of heat, and it may therefore be regarded as a generalization of that equation. The following account is in fact based upon that of Goursat (1923) for the heat equation.

We note also, for future reference, that the local rate of diffusion across $z = constant$ is given by

$$-A_z\frac{\partial\chi}{\partial z} = -Bz'^{1-2p}\frac{\partial\chi}{\partial z'}, \tag{1·7}$$

where

$$B = \frac{a'^{1-p}u_1^{1-2p}}{(2-4p)^{1-2p}z_1^m x_0^p}. \tag{1·8}$$

2. Boundary conditions

We begin by determining sets of boundary conditions which define a unique solution of the equation (1·6). Discarding the primes we write the equation as

$$\frac{\partial^2\chi}{\partial z^2} + \frac{1-2p}{z}\frac{\partial\chi}{\partial z} = \frac{\partial\chi}{\partial x},$$

where z, x are real dimensionless co-ordinates, z is positive and $0 < p < 1$. This may be written as $L(\chi) = 0$, where

$$L(U) \equiv \frac{\partial}{\partial z}\left(z^{1-2p}\frac{\partial U}{\partial z}\right) - z^{1-2p}\frac{\partial U}{\partial x}.$$

* O. G. Sutton (1934). It is easily verified that if, as in the evaporation theory, $a = N\nu^n z_1^{m-n}$, where N is a pure number and ν is the kinematic viscosity of the fluid, then x', z' are dimensionless variables.

The adjoint expression is

$$M(V) \equiv \frac{\partial}{\partial z}\left\{z^{1-2p}\frac{\partial V}{\partial z}\right\} + z^{1-2p}\frac{\partial V}{\partial x},$$

and we have, identically,

$$VL(U) - UM(V) \equiv \frac{\partial H}{\partial z} + \frac{\partial K}{\partial x},$$

where $$H = z^{1-2p}\left(V\frac{\partial U}{\partial z} - U\frac{\partial V}{\partial z}\right), \quad K = -z^{1-2p}UV.$$

Hence if C be the boundary of a region D with $z \geqslant 0$ throughout $D+C$, and if U, V are such that H, K and their derivatives are continuous in $D+C$, then we have Green's formula

$$\iint_D \{VL(U) - UM(V)\}\,dz\,dx + \int_C z^{1-2p}\left\{UV\,dz + \left(V\frac{\partial U}{\partial z} - U\frac{\partial V}{\partial z}\right)dx\right\} = 0, \quad (2{\cdot}2)$$

where the integral along C is taken in the direction from positive x to positive z. If in this we write $V = 1$ and $U = \chi^2$, where χ is a solution of (2·1) regular* in $D+C$, we obtain

$$2\iint_D z^{1-2p}\left(\frac{\partial \chi}{\partial z}\right)^2 dz\,dx + \int_C z^{1-2p}\left(\chi^2 dz + 2\chi\frac{\partial \chi}{\partial z}dx\right) = 0. \quad (2{\cdot}3)$$

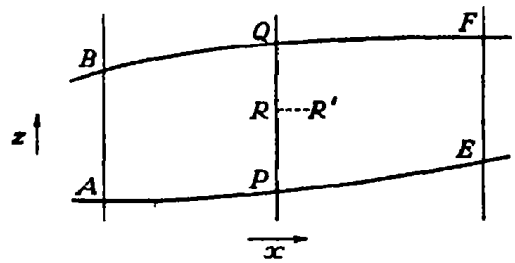


FIGURE 1

In figure 1, let AB, EF be two given characteristics $x = constant$ and PQ any intermediate characteristic. Let APE, BQF be two curves, along which $z \geqslant 0$, which meet any characteristic only once. Then we prove that there cannot be more than one solution of (2·1), regular inside and on the boundary of the region $ABFE$, which assumes† a given set of continuous values on AB, AE and BF. For, applying (2·3) to the region $ABQP$ and the difference $\bar{\chi}$ of two such solutions we get

$$2\iint_{ABQP} z^{1-2p}\left(\frac{\partial \bar{\chi}}{\partial z}\right)^2 dz\,dx + \int_{\overrightarrow{PQ}} z^{1-2p}\bar{\chi}^2 dz = 0, \quad (2{\cdot}4)$$

* i.e. such that χ and each term of $L(\chi)$ are continuous.

† Here and elsewhere, limiting values as the boundary is approached from within are to be understood.

since $\bar{\chi} = 0$ along AB, AP and BQ. The elements of both integrals are essentially non-negative, hence $\bar{\chi} = 0$ at all points of PQ and is therefore identically zero throughout $ABFE$.

In problems of evaporation or forced convection it is frequently possible to take AE as $z = 0$ and BF as $z = \infty$. The initial state of the fluid stream is assumed to be known, and the magnitude of the mean velocity is taken to be such that no appreciable upwind diffusion occurs. The condition that χ assumes a given set of values on AB, now a semi-infinite segment and normal to AE, is therefore appropriate to such problems. The boundary condition on AE requires more consideration. If AE forms part of a free liquid or saturated surface from which evaporation into an air stream is taking place, then a natural assumption and one which appears to yield very satisfactory results is that the vapour pressure of the evaporating liquid assumes its saturation value at the temperature of the surface as the surface is approached (Pasquill 1943). This implies that the vapour concentration χ assumes a constant value χ_s all along AE. Again, if the boundary is a dry surface impermeable to vapour the boundary condition is, by (1·7),

$$\lim_{z \to +0} \left\{ z^{1-2p} \frac{\partial \chi}{\partial z} \right\} = 0.$$

Thus for the case of initially dry air travelling over a saturated strip $0 < x < 1$ followed by dry ground impermeable to vapour, appropriate boundary conditions, as far as the boundaries $x = 0$ and $z = 0$ are concerned, are

$$\left.\begin{aligned} &\lim_{x \to +0} \chi = 0 \quad (z > 0), \\ &\lim_{z \to +0} \chi = \chi_s \quad (0 < x \leqslant 1), \\ &\lim_{z \to +0} \left\{ z^{1-2p} \frac{\partial \chi}{\partial z} \right\} = 0 \quad (x > 1). \end{aligned}\right\} \tag{2·5}$$

It is perhaps worth noting that in these conditions the rate of evaporation is not prescribed (e.g. by an empirical law) at the evaporating surface but only over the ensuing impermeable dry ground where its value is unequivocal.* Its value over the evaporating surface will, of course, be furnished by the solution of the problem.

From the mathematical point of view these conditions may be made more general. We shall consider three types of problems relative to the region $z > 0$, $x > x_1$ (a suitable fixed value of x), namely, those where on $z = 0$ either (i) the value of χ is prescribed or (ii) the local rate of evaporation is prescribed or (iii) the local rate of evaporation is given as a linear function of χ. As in (2·5) different conditions of these types may hold over different parts of the boundary, and in each case it is assumed

* In this respect the treatment differs from that of Giblett (1921) who uses as the boundary condition an empirical law, due to Bigelow, connecting the rate of evaporation with the vapour pressure and wind velocity.

in addition that the value of χ on $x = x_1$ $(z > 0)$ is prescribed. The boundary conditions defining these problems are therefore of the forms

$$\left.\begin{aligned} &\text{(i)} \quad \operatorname*{Lim}_{z\to 0} \chi = \psi(x), \\ &\text{(ii)} \quad \operatorname*{Lim}_{z\to 0} \left\{ -z^{1-2p} \frac{\partial \chi}{\partial z} \right\} = \phi(x), \\ &\text{(iii)} \quad \operatorname*{Lim}_{z\to 0} \left\{ z^{1-2p} \frac{\partial \chi}{\partial z} - h(x)\chi \right\} = g(x), \end{aligned}\right\} \quad (x > x_1) \tag{2·6}$$

with in each case the 'initial' condition

$$\operatorname*{Lim}_{x\to x_1+0} \chi = f(z) \quad (z > 0). \tag{2·7}$$

Here ψ, ϕ, h, g and f denote given functions.

To obtain uniqueness theorems for these problems some restriction must be imposed on the behaviour of χ as $z \to \infty$, in place of the assumption that χ is known along BF. Now from (2·3) it is seen that the contribution from BQ to the integral around $APQBA$ will vanish in the limit when BF is at infinity provided that

$$\operatorname*{Lim}_{z\to\infty} \left\{ z^{1-2p} \chi \frac{\partial \chi}{\partial z} \right\} = 0. \tag{2·8}$$

A solution satisfying this condition may be called *regular at* $z = \infty$. It is then easily verified that in cases (i) and (ii) there is at most one regular solution whose values $f(z)$ along AB themselves satisfy (2·8). In case (iii) we note that the difference $\bar{\chi}$ of two such solutions is zero along AB and along AE satisfies the condition

$$\operatorname*{Lim}_{z\to 0} \left\{ z^{1-2p} \frac{\partial \bar{\chi}}{\partial z} \right\} = \operatorname*{Lim}_{z\to 0} \{h\bar{\chi}\}.$$

Hence, applying (2·3) to $\bar{\chi}$ we reproduce (2·4) with the addition of a term $2\int_{AP} h\bar{\chi}^2 dx$ to the left-hand side. The solution is therefore unique provided that $h(x) \geqslant 0$.

3. Basic solutions

As in the case of the heat equation (Goursat 1923) we proceed to obtain certain 'basic solutions' of (2·1) from which integrals satisfying the boundary conditions of § 2 may be constructed. It is convenient, however, to obtain two such solutions in place of the single one which is usually employed for the heat equation.

The substitution $\chi = z^p \Omega$ reduces (2·1) to

$$\frac{\partial^2 \Omega}{\partial z^2} + \frac{1}{z}\frac{\partial \Omega}{\partial z} - \frac{p^2}{z^2}\Omega = \frac{\partial \Omega}{\partial x}, \tag{3·1}$$

which has solutions $\qquad e^{-\sigma^2(x-\mu)} J_{\pm p}(\sigma z),$

where σ, μ are arbitrary constants. Now with $0<p<1$ and $t>0$ we have (Watson 1922)

$$\int_0^\infty e^{-\sigma^2 t} J_{\pm p}(\sigma z) J_{\pm p}(\sigma\lambda)\,\sigma\,d\sigma = \frac{1}{2t} e^{-\frac{1}{4}(z^2+\lambda^2)/t} I_{\pm p}\left(\frac{\lambda z}{2t}\right).$$

Hence with $\lambda>0$, $z>0$ and $x>\mu$ we write $x-\mu$ for t and obtain the basic solutions of (2·1) as

$$\begin{aligned}\chi_{1,2}(z,x;\lambda,\mu) &= \frac{\lambda^p z^p}{2(x-\mu)^p} e^{-\frac{1}{4}(z^2+\lambda^2)/(x-\mu)} I_{\pm p}\left\{\frac{\lambda z}{2(x-\mu)}\right\} \\ &= (2x-2\mu)^{p-1} e^{-\frac{1}{4}(z^2+\lambda^2)/(x-\mu)} \omega^p I_{\pm p}(\omega), \qquad (3\cdot2)\end{aligned}$$

where the suffix 1 corresponds to the positive sign with p and

$$\omega = \tfrac{1}{2}\lambda z/(x-\mu).$$

As functions of λ, μ these solutions satisfy the equation adjoint to (2·1):

$$\frac{\partial}{\partial\lambda}\left(\lambda^{1-2p}\frac{\partial\chi}{\partial\lambda}\right) + \lambda^{1-2p}\frac{\partial\chi}{\partial\mu} = 0 \quad (z,\lambda>0;\ \mu<x).$$

They may be obtained rather more simply from the result that (3·1) is invariant under the transformation*

$$x' = -\frac{1}{x}, \quad z' = \frac{z}{x}, \quad \Omega = \frac{1}{x} e^{-\frac{1}{4}z^2/x}\Omega'.$$

For it follows that if $\Omega(z,x)$ is a solution, so is

$$\frac{1}{x-\mu} e^{-\frac{1}{4}z^2/(x-\mu)}\Omega\left(-\frac{1}{x-\mu}, \frac{z}{x-\mu}\right). \qquad (3\cdot3)$$

Applying this to the solutions $\Omega = e^{\frac{1}{2}\lambda^2 x} I_{\pm p}(\frac{1}{2}\lambda z)$ we obtain (3·2).

It may be noted that, since

$$I_{\frac{1}{2}}(\omega) = \left(\frac{2}{\pi\omega}\right)^{\frac{1}{2}}\sinh\omega, \quad I_{-\frac{1}{2}}(\omega) = \left(\frac{2}{\pi\omega}\right)^{\frac{1}{2}}\cosh\omega,$$

when $p=\frac{1}{2}$ we have $\quad \chi_1+\chi_2 = \dfrac{1}{\sqrt{\pi}} \dfrac{e^{-\frac{1}{4}(z-\lambda)^2/(x-\mu)}}{\sqrt{(x-\mu)}} \quad (x>\mu),$

which is the basic solution for the heat equation.

So far, $\chi_{1,2}$ have been defined only for $x>\mu$. To complete their definitions for $x\leqslant\mu$ we show that $\chi_{1,2}$ and all their z- or x-derivatives tend to zero as (z,x) approaches any point of $x=\mu$, $z>0$ from the right, other than the point (λ,μ). For, using the formula

$$\frac{d}{d\omega}\{\omega^{-\nu} I_\nu(\omega)\} = \omega^{-\nu} I_{\nu+1}(\omega),$$

* z', x' are of course distinct from the variables of § 1.

we see that any derivative is the sum of terms of the form

$$A\lambda^{p+n}z^b(x-\mu)^c e^{-\frac{1}{2}(z^2+\lambda^2)/(x-\mu)} I_{m\pm p}(\omega), \tag{3·4}$$

where m, n are non-negative integers. Now as $x \to \mu + 0$ with $z, \lambda > 0$ we have $\omega \to +\infty$. Hence, using the asymptotic expansion for $I_\nu(\omega)$,

$$I_\nu(\omega) = \frac{e^\omega}{\sqrt{(2\pi\omega)}}\{1 + O(\omega^{-1})\}, \tag{3·5}$$

we express $\chi_{1,2}$ and their derivates as the sum of terms like

$$B(x-\mu)^\alpha e^{-\frac{1}{2}(z-\lambda)^2/(x-\mu)},$$

where B is bounded as $x \to \mu + 0$. Hence with $|z-\lambda| > 0$ the result follows.

We therefore define $\chi_{1,2}$ to be zero for $x \leqslant \mu, z > 0$ As thus defined they are solutions of (2·1), regular for all $z, \lambda > 0$ and for all x, μ except at the point $z = \lambda, x = \mu$ where they are undefined.

We also require limiting forms of $\chi_{1,2}$ as $\lambda \to +0$. Considering $\lambda^{-2p}\chi_1$ and χ_2 and discarding a numerical factor we obtain the solutions, regular for $z > 0$ and all x,

$$\left.\begin{aligned} \chi_1(z, x; \mu) &= z^{2p}(x-\mu)^{-1-p} e^{-\frac{1}{2}z^2/(x-\mu)} \quad (x > \mu) \\ &= 0 \quad (x \leqslant \mu), \end{aligned}\right\} \tag{3·6}$$

$$\left.\begin{aligned} \chi_2(z, x, \mu) &= (x-\mu)^{p-1} e^{-\frac{1}{2}z^2/(x-\mu)} \quad (x > \mu) \\ &= 0 \quad (x \leqslant \mu). \end{aligned}\right\} \tag{3·7}$$

These solutions may be obtained directly by applying (3 3) to the solutions $\Omega = z^{\pm p}$ of (3·1).

From the basic solutions we construct the integrals we require for the solution of boundary value problems by integrating along suitable paths in the half-plane $\lambda \geqslant 0$. These will be of the general form

$$\chi(z, x) = \int_L (A\chi_1 + B\chi_2)(P\,d\lambda + Q\,d\mu), \tag{3·8}$$

where A, B, P, Q are functions of λ, μ. If L is a part of $\lambda = 0$ we shall of course replace $\chi_{1,2}$ by their limiting forms (3·6), (3·7) From what has been proved about $\chi_{1,2}$ it follows that in general the integral defined by (3·8) is regular throughout any bounded closed region in $z > 0$ which does not contain any part of L * Also since $\chi_{1,2}$ are zero for $\mu \geqslant x$ we need integrate only along that part of L for which $\mu < x$. Hence $\chi(z, x)$ defined by (3·8) is zero for all x less than the least value of μ on L and at all points not on L where x equals this least value. Applied to the physical problem, this means that there is no diffusion upwind, a result which it is reasonable to accept on physical grounds.

* The (z, x) and (λ, μ) planes are imagined to be superimposed.

4. Integrals satisfying initial conditions

If in (3·8) we take L as the semi-infinite segment $\lambda > 0$ of the characteristic $\mu = h$ we obtain integrals which enable us to satisfy the initial boundary condition (2·7) on this characteristic without disturbing the boundary condition over $z = 0$ in cases (i) and (ii) of (2·6).

We consider the expressions

$$F_{1,2}(z,x) = \int_0^\infty \lambda^{1-2p} f(\lambda)\, \chi_{1,2}(z,x;\lambda,h)\, d\lambda$$

$$= \int_0^\infty z^p \lambda^{1-p} f(\lambda)\, e^{-\frac{1}{2}(z^2+\lambda^2)/(x-h)} I_{\pm p}(\omega)\, \frac{d\lambda}{2(x-h)}, \tag{4·1}$$

$$(z>0,\ x>h,\quad 0<p<1;\quad \omega = \tfrac{1}{2}\lambda z/(x-h))$$

which are clearly formal solutions of (2·1). $F_{1,2}(z,x)$ are zero for $x<h$, $z>0$. We assume that $f(\lambda)$ is continuous for $\lambda \geqslant 0$, except possibly for a finite number of finite discontinuities, and that it is bounded as $\lambda \to +\infty$. It is readily seen that $F_{1,2}$ and the expressions obtained by formal differentiation under the integral sign any number of times with respect to z or x converge absolutely at any point $z>0$, $x>h$, and, moreover, uniformly throughout any bounded closed region $D+C$ contained in $z>0$, $x>h$. For, with $0<p<1$ and (3·4), convergence at $\lambda = 0$ is assured. Near $\lambda = \infty$ we use the asymptotic expansion (3·5) and, with $f(\lambda)$ bounded, obtain for the dominant terms in F or any derivative, expressions of the form

$$A(z,x,\mu)\, \lambda^n\, e^{-\frac{1}{2}(z-\lambda)^2/(x-h)},$$

where A is continuous in $D+C$. Hence the formal differentiations are valid and $F_{1,2}$ are integrals of (2·1), regular throughout $D+C$.

The fundamental property of F which enables us to satisfy the boundary condition over $x = h$ is

$$\left.\begin{aligned} \lim_{x\to h+0} F(z,x) &= \tfrac{1}{2}\{f(z-0)+f(z+0)\} \quad (z>0) \\ &= f(z), \end{aligned}\right\} \tag{4·2}$$

if $f(z)$ is continuous at z. To prove this we write in (4·1)

$$\int_0^\infty = \int_0^a + \int_a^\infty \quad (0<a<z). \tag{4·3}$$

In the first integral on the right we use the inequality *

$$0 < I_\nu(\omega) < (\tfrac{1}{2}\omega)^\nu\, e^\omega/\Gamma(1+\nu) \quad (\omega>0,\ \nu > -\tfrac{1}{2}),$$

* An immediate consequence of the formula (Whittaker & Watson 1920)

$$I_\nu(\omega) = \frac{(\frac{1}{2}\omega)^\nu}{\Gamma(\frac{1}{2})\,\Gamma(\nu+\frac{1}{2})} \int_0^\pi e^{\omega\cos\theta} \sin^{2\nu}\theta\, d\theta \quad (R(\nu) > -\tfrac{1}{2}).$$

for I_p and for I_{-p} with $p < \frac{1}{2}$. For I_{-p} with $\frac{1}{2} \leqslant p < 1$ we first use the recurrence formula

$$I_{-p}(\omega) = I_{2-p}(\omega) + 2(1-p)\, I_{1-p}(\omega)/\omega,$$

and apply the above inequality to each term on the right, getting

$$0 < I_{-p}(\omega) < (\tfrac{1}{2}\omega)^{-p}\, e^{\omega}/\Gamma(1-p) + (\tfrac{1}{2}\omega)^{2-p}\, e^{\omega}/\Gamma(3-p), \quad (\omega > 0).$$

We thus obtain in the case of F_1,

$$\left| \int_0^a \right| < \frac{z^{2p} e^{-\frac{1}{2}(z-a)^2/(x-h)}}{(2x-2h)^{1+p}\,\Gamma(1+p)} \int_0^a \lambda\, |f(\lambda)|\, d\lambda$$
$$< C\zeta^{2+2p}\, e^{-\frac{1}{2}(z-a)^2/(x-h)}, \tag{4·4}$$

where C is a positive constant and $\zeta = \frac{1}{2}z/\sqrt{(x-h)}$. In the case of F_2 the factor $C\zeta^{2+2p}$ in (4·4) is replaced by

$$C_1\zeta^{2-2p} + C_2\zeta^{6-2p}, \tag{4·4}'$$

where C_2 may be taken as zero if $0 < p < \frac{1}{2}$.

In the remaining integral in (4·3) the argument ω of the Bessel function tends to infinity through real positive values as $x \to h+0$, $z > 0$, hence we use (3·5) and obtain from the dominant term

$$I = \int_a^{\infty} \frac{f(\lambda)}{\sqrt{\pi}} \left(\frac{\lambda}{z}\right)^{\frac{1}{2}-p} e^{-\frac{1}{2}(z-\lambda)^2/(x-h)} \frac{d\lambda}{2\sqrt{(x-h)}}$$
$$= \int_a^{\infty} f(\lambda) \left(\frac{\lambda}{z}\right)^{\frac{1}{2}-p} H(z-\lambda, x-h)\, d\lambda, \tag{4·5}$$

where $$H(u, v) = \tfrac{1}{2} e^{-\frac{1}{4}u^2/v}/\sqrt{(\pi v)} \quad (v > 0).$$

We now write $a = z - \epsilon$, where $0 < \epsilon < z$. The limiting value of the expressions in (4·4), (4·4)′ as $x \to h+0$ is zero; that of I is obtained by a slight extension of a familiar argument. In fact we have from (4·5)

$$\left| I - f(z-0) \int_{z-\epsilon}^{z} H(z-\lambda, x-h)\, d\lambda - f(z+0) \int_{z}^{z+\epsilon} H(z-\lambda, x-h)\, d\lambda \right|$$
$$\leqslant \eta_1 \int_{z-\epsilon}^{z} H\, d\lambda + \eta_2 \int_{z}^{z+\epsilon} H\, d\lambda + M' \int_{z+\epsilon}^{\infty} (\lambda/z)^{\frac{1}{2}-p}\, H\, d\lambda,$$

where η_1, η_2 denote respectively the upper bounds of $|f(\lambda)(\lambda/z)^{\frac{1}{2}-p} - f(z \mp 0)|$ for $z-\epsilon \leqslant \lambda \leqslant z$ or $z \leqslant \lambda \leqslant z+\epsilon$, and M' is that of $|f(\lambda)|$ for $\lambda \geqslant z$. If now we write $\lambda = z + 2t\sqrt{(x-h)}$ and make $x \to h+0$ we find that

$$|I - \tfrac{1}{2}\{f(z-0) + f(z+0)\}| \leqslant \tfrac{1}{2}(\eta_1 + \eta_2), \tag{4·6}$$

since, for example,

$$\operatorname*{Lim}_{x \to h+0} \int_{z-\epsilon}^{z} H(z-\lambda, x-h)\, d\lambda = \operatorname*{Lim}_{x \to h+0} \frac{1}{\sqrt{\pi}} \int_{-\epsilon/\{2\sqrt{(x-h)}\}}^{0} e^{-t^2}\, dt = \tfrac{1}{2},$$

and the integral which is the coefficient of M' is convergent. But the left-hand side of (4·6) is independent of ϵ and the right-hand side is arbitrarily small with ϵ. Hence

$$\operatorname*{Lim}_{x\to h+0} I = \tfrac{1}{2}\{f(z-0)+f(z+0)\},$$

and (4·2) follows since the remainder of the asymptotic expansion (3·5) contributes nothing in the limit.

As a particular case we take $f(z)$ to be continuous for a finite positive range of values $z_1 \leqslant z \leqslant z_2$ and zero outside this range. Then we have

$$\operatorname*{Lim}_{x\to h+0} \int_{z_1}^{z_2} \lambda^{1-2p} f(\lambda)\, \chi_{1,2}(z, x, \lambda, h)\, d\lambda$$

equal to $f(z)$ if z is an interior point of the range (z_1, z_2), $\frac{1}{2}f(z)$ if $z = z_1$ or $z = z_2$ and zero if z is outside the range.

5 Behaviour of F as $z\to\infty$

In the preceding section it is assumed that z is neither zero nor infinite. The behaviour of the integrals (4·1) as $z\to\infty$ is given by the results

$$\left.\begin{aligned} \operatorname*{Lim}_{z\to\infty} F(z,x) &= \operatorname*{Lim}_{z\to\infty} f(z), && (5\cdot1)\\ \operatorname*{Lim}_{z\to\infty} \left\{z^{1-2p}\frac{\partial F}{\partial z}\right\} &= \operatorname*{Lim}_{z\to\infty} \{z^{1-2p} f'(z)\}, && (5\cdot2) \end{aligned}\right\} \quad (x>h)$$

provided that the limits on the right exist and $f'(z)$ is continuous for all sufficiently great values of z. Hence if $f(z)$ satisfies the condition (2·8), $F(z,x)$ will be regular at $z=\infty$ in the sense of § 2.

To prove these results we write $a = (1-\alpha)z$ in (4·3), where $0<\alpha<1$, and obtain from (4·4), for example,

$$\left|\int_0^a\right| < C\zeta^{2+2p}\, e^{-\frac{1}{4}\alpha^2 z^2/(x-h)}, \qquad (5\cdot3)$$

which tends to zero as $z\to\infty$, $x>h$. In the remaining integral we use the asymptotic expansion (3·5) as before to get (4·5). Writing in this $\lambda = z+2t\sqrt{(x-h)} = z(1+t/\zeta)$ we obtain

$$\begin{aligned} I &= \frac{1}{\sqrt{\pi}}\int_{-\alpha\zeta}^{\infty} f(\lambda)\,(1+t/\zeta)^{\frac{1}{2}-p}\, e^{-t^2}\, dt \\ &= \frac{1}{\sqrt{\pi}}\left(\int_{-\alpha\zeta}^{0} + \int_0^{\infty}\right). \end{aligned} \qquad (5\cdot4)$$

In the first of these integrals we use the binomial expansion of $(1+t/\zeta)^{\frac{1}{2}-p}$; in the second the inequality*

$$1 < (1+t/\zeta)^{\frac{1}{2}-p} < 1 + (\tfrac{1}{2}-p)\,(t/\zeta)\,(1+t/\zeta)^{\frac{1}{2}-p} \quad (t/\zeta>0),$$

* Cf. Whittaker & Watson (1920, 16·3).

if $0<p<\frac{1}{2}$, or $\quad 1-(p-\frac{1}{2})t/\zeta<(1+t/\zeta)^{\frac{1}{2}-p}<1$

if $\frac{1}{2}<p<1$. It follows that for large positive ζ

$$\left| I-\frac{1}{\sqrt{\pi}}\int_{-\alpha\zeta}^{\infty} f(\lambda)\,e^{-t^2}\,dt \right| < A/\zeta,$$

where A is a positive constant. Since $\lambda \geqslant (1-\alpha)z$ throughout the range of integration we find

$$\operatorname*{Lim}_{z\to\infty} I = \operatorname*{Lim}_{z\to\infty} f(z)$$

and (5·1) follows.

To prove (5·2) we have from (3·2) with $\mu = h$

$$\frac{\partial \chi_1}{\partial z} = \lambda^p z^p (2x-2h)^{-\frac{1}{2}} e^{-\frac{1}{2}(z^2+\lambda^2)/(x-h)} \{\lambda I_{p-1}(\omega) - z I_p(\omega)\}, \tag{5·5}$$

$$\frac{\partial \chi_2}{\partial z} = \lambda^p z^p (2x-2h)^{-\frac{1}{2}} e^{-\frac{1}{2}(z^2+\lambda^2)/(x-h)} \{\lambda I_{1-p}(\omega) - z I_{-p}(\omega)\}. \tag{5·6}$$

Using these in $z^{1-2p}\dfrac{\partial F}{\partial z}$ we obtain results like (5·3), (5·4) and deduce that, for large positive z,

$$z^{1-2p}\frac{\partial F}{\partial z} \sim \frac{1}{\sqrt{(2\pi)}}\int_{(1-\alpha)z}^{\infty} f(\lambda)\,(\lambda z)^{\frac{1}{2}-p}\,(\lambda - z)\,e^{-\frac{1}{2}(\lambda-z)^2/(x-h)}\,\frac{d\lambda}{(2x-2h)^{\frac{1}{2}}}$$

$$= -\frac{1}{\sqrt{\pi}}\int_{(1-\alpha)z}^{\infty} f(\lambda)\,(\lambda z)^{\frac{1}{2}-p}\,\frac{\partial}{\partial\lambda}\{e^{-\frac{1}{2}(\lambda-z)^2/(x-h)}\}\,\frac{d\lambda}{2\sqrt{(x-h)}}.$$

Integrating by parts we find that the integrated part tends to zero exponentially as $z\to\infty$, $x>h$. The remaining integral yields

$$\frac{1}{\sqrt{\pi}}\int_{-\alpha\zeta}^{\infty} \{(\tfrac{1}{2}-p)\lambda^{-2p} f(\lambda) + \lambda^{1-2p} f'(\lambda)\}\,(1+t/\zeta)^{p-\frac{1}{2}}\,e^{-t^2}\,dt.$$

The work is now as before, with obvious changes in the use of the inequalities for $(1+t/\zeta)^{\frac{1}{2}-p}$, and with $f(\lambda)$ bounded we get

$$\operatorname*{Lim}_{z\to\infty}\left\{z^{1-2p}\frac{\partial F}{\partial z}\right\} = \operatorname*{Lim}_{\lambda\to\infty}\{(\tfrac{1}{2}-p)\lambda^{-2p} f(\lambda) + \lambda^{1-2p} f'(\lambda)\}$$

$$= \operatorname*{Lim}_{\lambda\to\infty}\{\lambda^{1-2p} f'(\lambda)\},$$

which proves the result (5·2).

6. Behaviour of F as $z\to 0$ or as $x\to\infty$

As $z\to 0$ or $x\to\infty$ the argument of the Bessel function in (4·1) tends to zero. These cases may therefore conveniently be discussed together. If we use the power series

$$I_\nu(\omega) = \sum_{n=0}^{\infty} \frac{(\frac{1}{2}\omega)^{\nu+2n}}{\Gamma(n+1)\,\Gamma(n+\nu+1)},$$

and integrate term by term we obtain, formally,

$$F_1(z,x) = e^{-\zeta^2}\sum_{n=0}^{\infty} \frac{a_n(x)\,\zeta^{2n+2p}}{\Gamma(n+1)\,\Gamma(n+p+1)}, \tag{6·1}$$

where $\zeta = \frac{1}{2}z/\sqrt{(x-h)}$ as formerly and, with $\lambda^2 = 4s(x-h)$,

$$a_n(x) = \int_0^{\infty} f(2\sqrt{[s(x-h)]})\,e^{-s}s^n\,ds.$$

Hence $|a_n| < M\Gamma(n+1)$, where M is the upper bound of $|f(\lambda)|$ for $\lambda \geqslant 0$. It follows that (6·1) is dominated by

$$Me^{-\zeta^2}\sum_{n=0}^{\infty} \zeta^{2n+2p}/\Gamma(n+p+1), \tag{6·2}$$

and so converges absolutely for all ζ. Hence the term-by-term integration is valid,* and making $\zeta \to 0$ we find that $F_1(z,x)$ tends to zero like ζ^{2p} as $z \to 0$ or as $x \to \infty$. In fact

$$\operatorname*{Lim}_{z\to 0}\{\zeta^{-2p}F_1(z,x)\} = a_0(x)/\Gamma(1+p) \qquad (x>h), \tag{6·3}$$

$$\operatorname*{Lim}_{x\to\infty}\{\zeta^{-2p}F_1(z,x)\} = \operatorname*{Lim}_{\lambda\to\infty}\{f(\lambda)/\Gamma(1+p)\} \quad (z>0), \tag{6·4}$$

provided that this last limit exists. For if $\operatorname*{Lim}_{\lambda\to\infty} f(\lambda) = \beta$ then

$$|a_n(x) - \beta\Gamma(n+1)| < (M+|\beta|)\int_0^{(x-h)^{-\frac{1}{2}}} e^{-s}s^n\,ds + \epsilon\int_{(x-h)^{-\frac{1}{2}}}^{\infty} e^{-s}s^n\,ds,$$

where ϵ is the upper bound of $|f(\lambda)-\beta|$ for $\lambda \geqslant 2(x-h)^{\frac{1}{2}}$. The right-hand side is arbitrarily small for all sufficiently large values of x, hence $\operatorname*{Lim}_{x\to\infty} a_n(x) = \beta\Gamma(n+1)$. Putting $n = 0$ we deduce (6 4).

From (6·1) we have also

$$\operatorname*{Lim}_{z\to 0}\left\{z^{1-2p}\frac{\partial F_1}{\partial z}\right\} = 2^{1-2p}a_0(x)\,(x-h)^{-p}/\Gamma(p) \quad (x>h), \tag{6·5}$$

from which the local rate of absorption at the surface may be obtained.

If $f(z)$ is a constant M, $F_1(z,x)$ assumes a very simple form. For then $F_1(z,x)$ is given by the series (6·2) and we have

$$\frac{\partial F_1}{\partial \zeta^2} = Me^{-\zeta^2}\zeta^{2p-2}/\Gamma(p),$$

whence, since $F_1 = 0$ for $z = 0$, we deduce

$$F_1(z,x) = \frac{M}{\Gamma(p)}\int_0^{\zeta^2} e^{-t}t^{p-1}\,dt. \tag{6·6}$$

* Bromwich (1908, § 176).

In the case of $F_2(z, x)$ we obtain the series

$$F_2(z,x) = e^{-\zeta^2} \sum_{n=0}^{\infty} \frac{b_n(x)\,\zeta^{2n}}{\Gamma(n+1)\,\Gamma(n-p+1)}, \tag{6·7}$$

where

$$b_n(x) = \int_0^{\infty} f(2\sqrt{[s(x-h)]})\, e^{-s} s^{n-p}\, ds.$$

If $f(z)$ is a constant M, then $b_n(x) = M\Gamma(n-p+1)$ and (6·7) gives

$$F_2(z,x) = M. \tag{6·8}$$

In the general case we have

$$\left.\begin{aligned} \operatorname*{Lim}_{z\to 0} F_2(z,x) &= b_0(x) && (x>h),\\ \operatorname*{Lim}_{x\to\infty} F_2(z,x) &= \operatorname*{Lim}_{\lambda\to\infty} f(\lambda) && (z>0),\\ \text{and}\quad \operatorname*{Lim}_{z\to 0}\left\{z^{1-2p}\frac{\partial F_2}{\partial z}\right\} &= 0 && (x>h) \end{aligned}\right\} \tag{6·9}$$

The property that $\operatorname*{Lim}_{z\to 0} F_1(z,x) = 0$ makes F_1 suitable for use in boundary value problems of the type (i), as with it the initial condition over $x = h$ can be satisfied without disturbing the given boundary condition over $z = 0$. Similarly, (6·9) shows that $F_2(z,x)$ is the appropriate integral to use in order to satisfy the initial condition in problems of type (ii), where the local rate of evaporation is prescribed over $z = 0$. In the further discussion of these problems it may therefore be assumed that the initial condition to be satisfied is $\operatorname*{Lim}_{x\to h+0} \chi = 0$. Applied to the evaporation problem this means that the problem with air of a given 'initial' moisture content is reduced to one where the air stream is initially dry.

The integrals F_1 and F_2 may also be used in combination to satisfy both an initial condition and a boundary condition over $z = 0$. This is in fact done by O. G. Sutton (1934) who obtains for the vapour concentration above the evaporating strip an expression which is effectively $\chi = F_2 - F_1$, where the integrals are formed with $f(z) = \chi_s$, a constant. By (6·8) and (6·6) this is*

$$\chi(z,x) = \frac{\chi_s}{\Gamma(p)} \int_{\zeta^2}^{\infty} e^{-t} t^{p-1}\, dt. \tag{6·10}$$

In the general case, however, it is more convenient to use F_1 and F_2 to satisfy the initial condition and to deal with the boundary condition over $z = 0$ by means of the integrals which will be obtained in the next section.

No discussion has been given of the behaviour of $F(z,x)$ as (z,x) approaches one of the angular points A, B, E, F of figure 1 where $ABEF$ is the whole quadrant $z>0$, $x>h$. No difficulty arises at $z = \infty$, $x = h$, since the results (4·2) and (5·1) are in accord. But at the other points the limiting value of one or both of F_1, F_2 will

* Cf. Pasquill (1943, § 7).

depend upon the path by which the point is approached. In particular, it is clear from (6·6) that when $f(z)$ is a constant the limiting value of F_1 as $z \to 0$, $x \to h$ depends upon that of $z^2/(x-h)$.*

7. Integrals satisfying given boundary conditions on $z = 0$

To deal with boundary conditions of the types stated in § 2 we take the path L of (3·8) as a segment $a < \mu < b$ of $\lambda = 0$. From the general remark at the end of § 3 it follows that if $a < x < b$ we need integrate only from a to x.

(i) To solve problem (i) of (2·6) we consider

$$\Psi(z,x) = \frac{1}{2^{2p}\Gamma(p)} \int_a^b \psi(\mu)\, \chi_1(z,x;\mu)\, d\mu.$$

Hence
$$\Psi(z,x) = \frac{(\frac{1}{2}z)^{2p}}{\Gamma(p)} \int_a^x \psi(\mu)\, e^{-\zeta^2} (x-\mu)^{-1-p}\, d\mu \quad (a < x \leqslant b)$$

and
$$\Psi(z,x) = \frac{(\frac{1}{2}z)^{2p}}{\Gamma(p)} \int_a^b \psi(\mu)\, e^{-\zeta^2} (x-\mu)^{-1-p}\, d\mu \quad (x > b),$$

where, with a slight change of notation, $\zeta = \frac{1}{2}z/\sqrt{(x-\mu)}$. We assume that $\psi(x)$ is continuous for $a \leqslant x \leqslant b$. It is clear that Ψ is a solution of (2·1), regular for $x > a$, $z > 0$ and such that

$$\operatorname*{Lim}_{x \to a+0} \Psi = 0 \quad (z > 0).$$

The behaviour of Ψ as $z \to 0$ when $x > b$ offers no particular difficulty. In particular we have

$$\operatorname*{Lim}_{x \to 0} \Psi = 0 \quad (x > b). \tag{7·1}$$

Its behaviour as $z \to 0$, $a < x \leqslant b$ is given by the following theorem:

If $\psi(\mu)$ satisfies the Lipschitz condition

$$|\psi(\mu') - \psi(\mu)| < H\,|\mu' - \mu| \quad (a < \mu, \mu' \leqslant b).$$

Then for $a < x \leqslant b$
$$\operatorname*{Lim}_{z \to 0} \Psi(z,x) = \psi(x), \tag{7·2}$$

and
$$\operatorname*{Lim}_{z \to 0} \left\{ -z^{1-2p} \frac{\partial \Psi}{\partial z} \right\} = \frac{2^{1-2p}}{\Gamma(p)} \left\{ \frac{\psi(x)}{(x-a)^p} + p \int_a^x \frac{\psi(x) - \psi(\mu)}{(x-\mu)^{1+p}}\, d\mu \right\}. \tag{7·3}$$

If $\psi'(\mu)$ exists and is continuous for $a \leqslant x \leqslant b$ the latter result may be replaced by

$$\operatorname*{Lim}_{z \to 0} \left\{ -z^{1-2p} \frac{\partial \Psi}{\partial z} \right\} = \frac{2^{1-2p}}{\Gamma(p)} \frac{d}{dx} \int_a^x \frac{\psi(\mu)}{(x-\mu)^p}\, d\mu. \tag{7·4}$$

* Cf. Goursat (1923, p. 321, Ex. 3).

To prove these results we write $\Psi = \Psi_0 + \Psi_1$, where, for $a < x \leqslant b$,

$$\Psi_0(z,x) = \frac{(\frac{1}{2}z)^{2p}}{\Gamma(p)}\psi(x)\int_a^x e^{-\zeta^2}(x-\mu)^{-1-p}\,d\mu,$$

$$\Psi_1(z,x) = \frac{(\frac{1}{2}z)^{2p}}{\Gamma(p)}\int_a^x \frac{\psi(\mu)-\psi(x)}{(x-\mu)^{1+p}}\,e^{-\zeta^2}\,d\mu.$$

In Ψ_0 we make the substitution $\zeta^2 = t$ and obtain

$$\Psi_0(z,x) = \frac{\psi(x)}{\Gamma(p)}\int_{\frac{1}{4}z^2/(x-a)}^{\infty} e^{-t}t^{p-1}\,dt. \tag{7·5}$$

It follows immediately that

$$\left.\begin{aligned} &\lim_{z\to 0}\Psi_0 = \psi(x), \\ \text{and}\quad &\lim_{z\to 0}\left\{-z^{1-2p}\frac{\partial\Psi_0}{\partial z}\right\} = \frac{2^{1-2p}\psi(x)}{\Gamma(p)(x-a)^p}. \end{aligned}\right\} \quad (a<x\leqslant b) \tag{7·6}$$

Again, using the Lipschitz condition we obtain

$$|\Psi_1| < \frac{H(\frac{1}{2}z)^{2p}}{\Gamma(p)}\int_a^x (x-\mu)^{-p}\,d\mu.$$

Hence $\lim_{z\to 0}\Psi_1 = 0$ and (7·2) follows. Now

$$-z^{1-2p}\frac{\partial\Psi_1}{\partial z} = \frac{2^{1-2p}p}{\Gamma(p)}\int_a^x \frac{\psi(x)-\psi(\mu)}{(x-\mu)^{1+p}}\,e^{-\zeta^2}\,d\mu - \frac{z^2}{2^{1+2p}\Gamma(p)}\int_a^x \frac{\psi(x)-\psi(\mu)}{(x-\mu)^{2+p}}\,e^{-\zeta^2}\,d\mu. \tag{7·7}$$

The first integral is a continuous* function of (z,x) for $z \geqslant 0$, $a < x \leqslant b$, hence its limiting value as $z \to 0$ is obtained by putting $\zeta = 0$. Using the Lipschitz condition and the substitution $\zeta^2 = t$ we easily show that the second term on the right of (7·7) is numerically less than $\frac{1}{2}Hz^{2-2p}e^{-\frac{1}{4}z^2/(x-a)}$, and so, with $p < 1$, tends to zero with z. Hence

$$\lim_{z\to 0}\left\{-z^{1-2p}\frac{\partial\Psi_1}{\partial z}\right\} = \frac{2^{1-2p}p}{\Gamma(p)}\int_a^x \frac{\psi(x)-\psi(\mu)}{(x-\mu)^{1+p}}\,d\mu,$$

and from this with (7·6) we obtain (7·3).

If $\psi'(\mu)$ exists and is continuous in (a,b) then, integrating by parts, we have

$$p\int_a^x \frac{\psi(x)-\psi(\mu)}{(x-\mu)^{1+p}}\,d\mu = \frac{\psi(a)-\psi(x)}{(x-a)^p} + \int_a^x \frac{\psi'(\mu)\,d\mu}{(x-\mu)^p}.$$

Hence (7·3) yields

$$\begin{aligned}\lim_{z\to 0}\left\{-z^{1-2p}\frac{\partial\Psi}{\partial z}\right\} &= \frac{2^{1-2p}}{\Gamma(p)}\left\{\frac{\psi(a)}{(x-a)^p} + \int_a^x \frac{\psi'(\mu)\,d\mu}{(x-\mu)^p}\right\} \\ &= \frac{2^{1-2p}}{\Gamma(p)}\frac{d}{dx}\int_a^x \frac{\psi(\mu)\,d\mu}{(x-\mu)^p}\end{aligned}$$

by a known result. This is (7·4).

* Goursat (1923, § 504).

These results may also be proved from the equation, obtained by writing $\zeta^2 = t$,

$$\Psi(z,x) = \frac{1}{\Gamma(p)}\int_{\frac{1}{2}z^2/(x-a)}^{\infty} \psi(x-\tfrac{1}{2}z^2/t)\,e^{-t}t^{p-1}\,dt \quad (a<x\leqslant b). \tag{7·8}$$

From this it is clear also that

$$\operatorname*{Lim}_{z\to\infty}\Psi = 0 \quad (a<x).$$

Also from (7·5) and (7·7) we see that

$$\operatorname*{Lim}_{z\to\infty}\left\{-z^{1-2p}\frac{\partial\Psi}{\partial z}\right\} = 0,$$

in virtue of the exponential factor $e^{-\frac{1}{2}z^2/(x-a)}$. Hence Ψ is regular at $z=\infty$, in the sense of § 2.

It will be observed from (7·5) and (6·10) that if $\psi(x)$ and $f(z)$ are both equal to the same constant χ_s and we write $h=a$, then $\Psi = F_2 - F_1$ $(a<x\leqslant b)$. In fact these functions are each the solution of the same boundary value problem, both being zero on $x=a$, and equal to the same value χ_s on $z=0$, $a<x\leqslant b$.

The result (7·4) is particularly convenient for the calculation of the total rate of evaporation E when the surface concentration of vapour over the strip $a<x<b$ is $\psi(x)$. For, by (1·7) and (1·5),

$$E = Bx_0\int_a^b\left\{-z^{1-2p}\frac{\partial\Psi}{\partial z}\right\}_{z=0}dx$$

$$= \frac{Bx_0 2^{1-2p}}{\Gamma(p)}\int_a^b \frac{\psi(\mu)\,d\mu}{(b-\mu)^p}$$

by (7·4) If in this we write $\psi(\mu)=\chi_s$, the (constant) saturation value of χ at the temperature considered, we obtain the saturation value E_s of E. Hence

$$\frac{E}{E_s} = \frac{1-p}{(b-a)^{1-p}\chi_s}\int_a^b \frac{\psi(\mu)\,d\mu}{(b-\mu)^p}, \tag{7·9}$$

where, of course, the rates of evaporation are those per unit length crosswind.

It is perhaps necessary to observe here that since (7·9) is obtained under the assumption that conditions are steady, the result is not applicable without further investigation to the case where the non-uniform concentration $\psi(x)$ arises from a drying surface. For the vapour concentration χ will then involve the time and is in fact the solution of an equation obtained from (1·1) by the addition of a term $\partial\chi/\partial t$ to the left-hand side. The solution of this non-steady equation, however, is so complicated that (7·9) has been used with some success to provide a first approximation for the instantaneous rate of evaporation when the drying is slow.*

* K. L. Calder, W. G. L. Sutton (unpublished). $\psi(x)$ must of course be replaced by its instantaneous value $\psi(x,t)$. See also § 10 of this paper.

(ii) For problem (ii) where the rate of evaporation is prescribed over a segment of $z = 0$ we consider the integral

$$\Phi(z,x) = \frac{2^{2p-1}}{\Gamma(1-p)}\int_a^b \phi(\mu)\,\chi_2(z,x,\mu)\,d\mu.$$

Hence
$$\Phi(z,x) = \frac{2^{2p-1}}{\Gamma(1-p)}\int_a^x \phi(\mu)\,(x-\mu)^{p-1}e^{-\zeta^2}\,d\mu \quad (a<x\leqslant b)$$

and
$$\Phi(z,x) = \frac{2^{2p-1}}{\Gamma(1-p)}\int_a^b \phi(\mu)\,(x-\mu)^{p-1}e^{-\zeta^2}\,d\mu \quad (x>b).$$

It is clear that $\operatorname{Lim}_{x\to a+0}\Phi = 0 \quad (z>0), \qquad \operatorname{Lim}_{z\to\infty}\Phi = 0 \quad (a<x);$

and
$$\operatorname{Lim}_{z\to 0}\Phi = \frac{2^{2p-1}}{\Gamma(1-p)}\int_a^x \phi(\mu)\,(x-\mu)^{p-1}\,d\mu \quad (a<x\leqslant b). \tag{7·10}$$

To obtain the rate of evaporation we form

$$-z^{1-2p}\frac{\partial\Phi}{\partial z} = \frac{(\frac{1}{2}z)^{2-2p}}{\Gamma(1-p)}\int_a^x \phi(\mu)\,(x-\mu)^{p-2}e^{-\zeta^2}\,d\mu \quad (a<x\leqslant b), \tag{7·11}$$

and a similar expression with the upper limit replaced by b when $x>b$ Now if on the right of (7·11) we write $1-p$ for p we obtain an expression of the form of $\Psi(z,x)$.* Hence from (7 2) and (7·1) we deduce at once

$$\operatorname{Lim}_{z\to 0}\left\{-z^{1-2p}\frac{\partial\Phi}{\partial z}\right\} = \begin{matrix}\phi(x) & (a<x\leqslant b),\\ 0 & (x>b)\end{matrix} \tag{7·12}$$

Also, from the behaviour of Ψ at $z=\infty$, we have

$$\operatorname{Lim}_{z\to\infty}\left\{z^{1-2p}\frac{\partial\Phi}{\partial z}\right\} = 0 \quad (a<x),$$

so that Φ is regular at $z=\infty$ in the sense of § 2. Hence $\Phi(z,x)$ is the solution of problem (ii) with $\chi = 0$ over $x=a$ From (7 12) it is seen that $\Phi(z,x)$ will also give the vapour concentration over the impermeable dry ground beyond an evaporating strip provided that $\phi(x)$ is chosen so as to satisfy the condition $\operatorname{Lim}_{z\to 0}\Phi = \psi(x)$ over the strip $a<x<b$ From (7·10) this condition is

$$\psi(x) = \frac{2^{2p-1}}{\Gamma(1-p)}\int_a^x \phi(\mu)\,(x-\mu)^{p-1}\,d\mu \quad (a<x\leqslant b). \tag{7·13}$$

Alternatively, from (7·4) and the uniqueness of solution of the second boundary value problem we obtain the condition as

$$\phi(x) = \frac{2^{1-2p}}{\Gamma(p)}\frac{d}{dx}\int_a^x \psi(\mu)\,(x-\mu)^{-p}\,d\mu \quad (a<x\leqslant b). \tag{7·14}$$

That these conditions are equivalent follows from the known solution of Abel's integral equation.

* The general theorem of which this is a particular case is given in § 8.

We may now complete the solution of the problem of the evaporation from an isolated saturated strip, defined by conditions (2·5), by determining the vapour concentration at any point downwind beyond the strip. For if $\psi(x) = \chi_s$, the constant saturation value of χ, then by (7·14) we have

$$\phi(x) = \frac{2^{1-2p}\chi_s}{(x-a)^p\,\Gamma(p)},$$

and the complete solution of problem (2·5) appears in the form

$$\chi(z,x) = \frac{\chi_s}{\Gamma(p)\,\Gamma(1-p)}\int_0^x \mu^{-p}(x-\mu)^{p-1}e^{-\zeta^2}d\mu \quad (0<x\leqslant 1),$$

$$\chi(z,x) = \frac{\chi_s}{\Gamma(p)\,\Gamma(1-p)}\int_0^1 \mu^{-p}(x-\mu)^{p-1}e^{-\zeta^2}d\mu \quad (x>1). \tag{7·15}$$

The first of these expressions is readily reduced to the form (6·10). These solutions have also been obtained by Calder, using the method of sources.

(iii) The solution of problem (iii) of (2·6) where the boundary condition involves both χ and $z^{1-2p}\dfrac{\partial\chi}{\partial z}$ is not so immediate as that of problems (i) and (ii). We suppose that the problem is first reduced to one with $\chi = 0$ on AB by means of F_1 or F_2, and we assume without loss of generality that AE is $z = 0$, $0<x<1$ on which the boundary condition is

$$\operatorname*{Lim}_{z\to 0}\left\{z^{1-2p}\frac{\partial\chi}{\partial z} - h(x)\,\chi\right\} = g(x),$$

where $h(x) > 0$. If we seek to satisfy this condition by an integral of the type Φ we obtain for $\phi(x)$ the Volterra integral equation

$$\phi(x) + \frac{2^{2p-1}}{\Gamma(1-p)}h(x)\int_0^x \phi(\mu)\,(x-\mu)^{p-1}d\mu + g(x) = 0 \quad (0<x\leqslant 1),$$

whose solution can be obtained (see, for example, Volterra & Pérès 1936).

Let us consider a case where $h(x)$ and $g(x)$ are constants, namely, that where the boundary condition is

$$\operatorname*{Lim}_{z\to 0}\left\{z^{1-2p}\frac{\partial\chi}{\partial z} + \lambda(\chi_s-\chi)\right\} = 0 \quad (0<x\leqslant 1), \tag{7·16}$$

χ_s and λ being positive constants. With a slight change of notation we write

$$\chi(z,x) = \frac{\lambda\chi_s}{2^{1-2p}\Gamma(1-p)}\int_0^x \phi(\mu)\,(x-\mu)^{p-1}e^{-\frac{1}{4}z^2/(x-\mu)}d\mu,$$

and obtain for $\phi(x)$ the integral equation

$$\phi(x) + \frac{\rho}{\Gamma(p)}\int_0^x \phi(\mu)\,(x-\mu)^{p-1}d\mu = 1,$$

where

$$\rho = 2^{2p-1}\lambda\,\Gamma(p)/\Gamma(1-p).$$

The solution is (Volterra & Pérès 1936)

$$\phi(x) = \sum_{r=0}^{\infty} (-\rho x^{\nu})^r / \Gamma(rp+1), \tag{7·17}$$

a series which converges absolutely for all values of x.

If $p = 1/\nu$, where ν is a positive integer, we may obtain another expression for $\phi(x)$ as follows. We write $\rho^{\nu} x = \xi$ and rearrange (7·17) by grouping together the terms separated by intervals of ν. Thus we get

$$\phi(x) = \sum_{s=0}^{\nu-1} \phi_s(x),$$

where $$\phi_s(x) = (-1)^s \left\{ \frac{\xi^{sp}}{\Gamma(sp+1)} \pm \frac{\xi^{sp+1}}{\Gamma(sp+2)} + \frac{\xi^{sp+2}}{\Gamma(sp+3)} \pm \ldots \right\},$$

and the upper or lower signs are taken according as ν is even or odd. Then, noting that $\phi_0(x) = e^{\pm\xi}$, we get

$$\frac{\partial \phi}{\partial \xi} \mp \phi = \sum_{s=1}^{\nu-1} (-1)^s \xi^{sp-1} / \Gamma(sp),$$

whence on integration with the conditon $\phi = 1$ for $\xi = 0$ we obtain

$$\phi(x) = \begin{cases} e^{\xi} \left\{ 1 + \sum_{s=1}^{\nu-1} \frac{(-1)^s}{\Gamma(sp)} \int_0^{\xi} e^{-t} t^{s\nu-1} \, dt \right\} & (\nu \text{ even}), \\ e^{-\xi} \left\{ 1 + \sum_{s=1}^{\nu-1} \frac{(-1)^s}{\Gamma(sp)} \int_0^{\xi} e^{t} t^{s\nu-1} \, dt \right\} & (\nu \text{ odd}) \end{cases}$$

In particular, if $p = \frac{1}{2}$ we have $\rho = \lambda$, $\xi = \lambda^2 x$ and

$$\phi(x) = e^{\lambda^2 x} \left\{ 1 - \frac{1}{\sqrt{\pi}} \int_0^{\lambda^2 x} e^{-t} t^{-\frac{1}{2}} \, dt \right\}$$

$$= \frac{e^{\lambda^2 x}}{\sqrt{\pi}} \int_{\lambda^2 x}^{\infty} e^{-t} t^{-\frac{1}{2}} \, dt.$$

From this we may obtain the usual form of the solution of a problem in heat conduction (Riemann-Weber 1901) In the general case we have

$$\operatorname*{Lim}_{z \to 0} \left\{ -z^{1-2p} \frac{\partial \chi}{\partial z} \right\} = \lambda \chi_s \phi(x),$$

whence the total rate of evaporation E is given by

$$\frac{E}{E_s} = \rho \Gamma(2-p) \int_0^1 \phi(x) \, dx$$

$$= \rho \Gamma(2-p) \sum_{r=0}^{\infty} (-\rho)^r / \Gamma(rp+2), \tag{7·18}$$

where, as previously, E_s denotes the total rate of evaporation from a saturated strip of the same length under the same external conditions of temperature, etc.

The boundary condition (7·16) is of interest in that it includes as limiting cases the condition $\chi = \chi_s$ assumed to hold over a saturated surface, and the condition of zero rate of evaporation appropriate to dry ground. These correspond to the cases $\lambda = \infty$ and $\lambda = 0$ respectively. It is therefore conceivable that a condition of this form, with λ a suitably chosen function of the time, may hold in certain cases throughout the whole of the drying process. The solution just obtained may in that case yield an approximate result when the surface is nearly dry, i.e. when λ is small.

8. Integrals applicable to other regions

In general, when the bounding curves AE, BF of figure 1 are not $z = 0$ and $z = \infty$, more general integrals must be constructed by using the basic solutions (3·2) instead of their limiting forms We suppose, therefore, that the path of integration L of (3·8) is the segment of the curve $\lambda = \theta(\mu)$ between $\mu = a$ and $\mu = b$. We assume that $\theta(\mu) > 0$ and that $\theta'(\mu)$ is continuous in (a, b). Considering

$$\int_a^b \phi(\mu)\, \chi_1(z, x, \theta(\mu), \mu)\, d\mu$$

we obtain the integral

$$\Phi_1(z, x) = \int_a^x [\phi(\mu)\{z\theta(\mu)\}^{\nu-\frac{1}{2}}(2\omega)^{\frac{1}{2}} e^{-\omega} I_\nu(\omega)]\, e^{-\frac{1}{4}\{z-\theta(\mu)\}^2/(x-\mu)} \frac{d\mu}{2\sqrt{(x-\mu)}} \qquad (8·1)$$

for $a < x \leqslant b$, and a similar expression with the upper limit replaced by b for $x > b$. Here

$$2\omega = z\theta(\mu)/(x-\mu),$$

and $\phi(\mu)$ is assumed to satisfy Lipschitz's condition in (a, b). Using the asymptotic expansion of $I_\nu(\omega)$ we see that the expression in square brackets is bounded for (z, x) on L. It follows that Φ_1 is continuous in the neighbourhood of L so that, as (z, x) approaches a point (Z, X) of L,

$$\operatorname{Lim} \Phi_1(z, x) = \Phi_1(Z, X).$$

In particular, if L is the straight line $z = Z$ we have $\theta(\mu) = Z$, and, for $a < X < b$,

$$\operatorname{Lim} \Phi_1 = \tfrac{1}{2} Z^{2\nu} \int_a^X \phi(\mu)\, e^{-\frac{1}{2}Z^2/(X-\mu)} I_\nu\left\{\frac{Z^2}{2(X-\mu)}\right\} \frac{d\mu}{X-\mu}. \qquad (8·2)$$

To discuss $z^{1-2\nu} \dfrac{\partial \Phi_1}{\partial z}$ as (z, x) approaches a point of L we use (5·5) and obtain

$$-z^{1-2\nu} \frac{\partial \Phi_1}{\partial z} = \int_a^x f_1(z, x, \mu)\, e^{-\frac{1}{4}\{z-\theta(\mu)\}^2/(x-\mu)} (x-\mu)^{-\frac{1}{2}}\, d\mu, \qquad (8·3)$$

where

$$f_1(z, x, \mu) = \tfrac{1}{4}\phi(\mu)\{z/\theta(\mu)\}^{\frac{1}{2}-\nu} (2\omega)^{\frac{1}{2}} e^{-\omega}\{z I_\nu(\omega) - \theta(\mu)\, I_{\nu-1}(\omega)\}. \qquad (8·4)$$

Using the asymptotic expansion for the Bessel functions we write (8·3) as

$$-z^{1-2p}\frac{\partial \Phi_1}{\partial z} = \frac{1}{4\sqrt{\pi}}\int_a^x \phi(\mu)\left\{\frac{z}{\theta(\mu)}\right\}^{\frac{1}{2}-p}\frac{z-\theta(\mu)}{(x-\mu)^{\frac{1}{2}}}e^{-\frac{1}{4}(z-\theta(\mu))^2/(x-\mu)}\,d\mu + I,$$

where I is of the form $\displaystyle\int_a^x H(z,x;\mu)\,(x-\mu)^{-\frac{1}{2}}\,d\mu,$

with H bounded for (z,x) on L, $a \leqslant \mu \leqslant b$. Hence I is continuous in the neighbourhood of L. The remaining integral is written as

$$\int_a^x \{\eta(z,\mu)-\eta(z,x)\}\frac{z-\theta(\mu)}{(x-\mu)^{\frac{1}{2}}}e^{-\frac{1}{4}(z-\theta(\mu))^2/(x-\mu)}\,d\mu + \eta(z,x)\int_a^x \frac{z-\theta(\mu)}{(x-\mu)^{\frac{1}{2}}}e^{-\frac{1}{4}(z-\theta(\mu))^2/(x-\mu)}\,d\mu,$$

where $$\eta(z,\mu) = \frac{1}{4\sqrt{\pi}}\phi(\mu)\{z/\theta(\mu)\}^{\frac{1}{2}-p}.$$

If $\phi(\mu)$ satisfies Lipschitz's condition in (a,b) the same is true of $\eta(z,\mu)$ as a function of μ, and the first term in the above expression is a continuous function of (z,x) in the neighbourhood of L. If the integral in the second part be denoted by $G(z,x)$ then it is known (Goursat 1923, § 544 (27)) that

$$\operatorname{Lim} G(z,x) = G(Z,X) \pm 2\sqrt{\pi},$$

where the sign to be taken is that of $z-Z$. Since also the limiting value of $\eta(z,x)$ is $\phi(X)$ we obtain finally

$$\operatorname{Lim}\left\{-z^{1-2p}\frac{\partial \Phi_1}{\partial z}\right\} = \int_a^X f_1(Z,X,\mu)e^{-\frac{1}{4}(Z-\theta(\mu))^2/(X-\mu)}\frac{d\mu}{(X-\mu)^{\frac{1}{2}}} \pm \tfrac{1}{2}\phi(X). \tag{8·5}$$

Precisely similar results are true for the integral $\Phi_2(z,x)$ formed from the basic solution χ_2 as Φ_1 was formed from χ_1. The analogue of (8·5) is obtained by replacing f_1 by f_2 where, from (5·6),

$$f_2(z,x;\mu) = \tfrac{1}{4}\phi(\mu)\{z/\theta(\mu)\}^{\frac{1}{2}-p}(2\omega)^{\frac{1}{2}}e^{-\omega}\{zI_{-p}(\omega)-\theta(\mu)I_{1-p}(\omega)\}. \tag{8·6}$$

In the particular case where L is the straight line $z = Z$ we have

$$\left.\begin{aligned} f_1(Z,X,\mu) &= \frac{Z^2}{4}\phi(\mu)(X-\mu)^{-\frac{1}{2}}e^{-\omega'}\{I_p(\omega')-I_{p-1}(\omega')\} \quad [\omega' = \tfrac{1}{2}Z^2/(X-\mu)],\\ f_2(Z,X;\mu) &= \frac{Z^2}{4}\phi(\mu)(X-\mu)^{-\frac{1}{2}}e^{-\omega'}\{I_{-p}(\omega')-I_{1-p}(\omega')\}, \end{aligned}\right\} \tag{8·7}$$

and the exponential factor in (8·5) and its analogue reduces to unity. In the special case $p = \frac{1}{2}$ we see that

$$f_1(Z,X,\mu)+f_2(Z,X,\mu) = 0,$$

so that here (Goursat 1923)

$$\operatorname*{Lim}_{z\to Z}\left\{\frac{\partial}{\partial z}(\Phi_1+\Phi_2)\right\} = \pm\phi(X).$$

In general, however, the integral term in (8·5) will not vanish, so that this last result is special for the heat equation.

It will be observed that, contrary to what might have been expected from § 7, both χ_1 and χ_2 generate integrals of the type Φ. To obtain integrals of the type Ψ we use the following theorem, which is easily verified:

If χ satisfies the equation (2·1) *then* $\chi' \equiv z^{1-2p}\dfrac{\partial \chi}{\partial z}$ *is a solution of the associated equation*

$$\frac{\partial^2 \chi'}{\partial z^2} - \frac{1-2p}{z}\frac{\partial \chi'}{\partial z} = \frac{\partial \chi'}{\partial x}, \tag{8·8}$$

obtained from (2·1) *by writing* $1-p$ *for* p.

With $0 < p < 1$ the relation is evidently a reciprocal one so that we have also:

If χ' satisfies (8·8) *then* $z^{2p-1}\dfrac{\partial \chi'}{\partial z}$ *is a solution of* (2·1).

For $p = \frac{1}{2}$ the theorems are trivial.

The first theorem has already been applied in § 7 (ii), the second, applied to integrals Φ' of (8·8) corresponding to (8·1), will yield integrals of (2·1) of the type Ψ. Thus if $f_1'(z, x; \mu)$ denotes the expression formed from (8·4) by writing $1-p$ for p and $\psi(\mu)$ for $\phi(\mu)$ the integral required is

$$\Psi_1(z, x) = \int_a^x f_1'(z, x; \mu)\, e^{-\frac{1}{4}(z-\theta(\mu))^2/(x-\mu)}\,(x-\mu)^{-1}\,d\mu, \tag{8·9}$$

with the property that as (z, x) tends to a point (Z, X) of L,

$$\operatorname{Lim} \Psi_1(z, x) = \Psi_1(Z, X) \pm \tfrac{1}{2}\psi(X).$$

Ψ_2 is obtained similarly from $f_2'(z, x; \mu)$.

If we apply the second theorem to the basic solution $\chi_2'(z, x, \mu)$ of (8·8) we obtain the solution $\chi_1(z, x; \mu)$ of (2·1). This mode of deriving $\chi_1(z, x, \mu)$ is less direct than the methods used in § 2 but shows why χ_1 generates an integral Ψ of a different type from Φ of § 7. No such relation exists between $\chi_1(z, x; \lambda, \mu)$ and $\chi_2'(z, x, \lambda, \mu)$.

The use of the integrals $\Phi_{1,2}$ and $\Psi_{1,2}$ in boundary value problems is sufficiently shown by that of their analogues for the heat equation (Goursat 1923, § 547). The initial condition will of course now be given over a finite range AB of figure 1 so that the integrals $F(z, x)$ will be formed with a finite range for λ.

9. Application to experiment

The results of the last section enable us to deal more fully with a point raised in § 1. The solution given by (6·10) for the evaporation from a saturated strip assumes that the power law (1·2) for the fluid velocity holds for unrestricted values of z, whereas under actual conditions, more especially those holding for wind-tunnel experiments, the range of its validity is restricted (Pasquill 1943, §§ 3, 5, 7, 13). It is of some importance, therefore, to obtain an idea of the order of magnitude of the correction to be made to the calculated rate of evaporation if the limited range of validity of the power law is taken into account.

We return to the notation of § 1, where z, x are the actual space co-ordinates and

z', x' are dimensionless variables given in terms of z, x by (1·5). In particular, we write $x' = x/x_0$, where x_0 is the length of the evaporating strip. Let us consider the simplified case where the power law for u and the resulting expression (1·3) for the diffusion coefficient A_z hold for $0 \leqslant z \leqslant c$. For $z > c$ we shall assume that u and A_z are constants, maintaining the values u_c, A_c they assume by continuity at $z = c$. At $z = c$ we assume also that χ and the flux $A_z \dfrac{\partial \chi}{\partial z}$ are continuous. Then if $z' = c'$ when $z = c$, χ must satisfy the conditions:

(i) $$\frac{\partial^2 \chi}{\partial z'^2} + \frac{1-2p}{z'} \frac{\partial \chi}{\partial z'} = \frac{\partial \chi}{\partial x'} \quad (0 < z' < c',\ 0 < x' < 1),$$

$$\lim_{z' \to 0} \chi = \chi_s \quad (0 < x' < 1),$$

(ii) $$\frac{\partial \chi}{\partial x} = \sigma^2 \frac{\partial^2 \chi}{\partial z^2} \quad \left(\sigma^2 = \frac{A_c}{u_c};\ z > c;\ 0 < x < x_0\right),$$

$$\lim_{z \to \infty} \chi = 0 \quad (0 < x < x_0),$$

(iii) χ is continuous at $z = c$ and

$$\lim_{z' \to c'-0} \left\{ B z'^{1-2p} \frac{\partial \chi}{\partial z'} \right\} = \lim_{z \to c+0} \left\{ A_c \frac{\partial \chi}{\partial z} \right\},$$

where B is the constant of (1·8),

(iv) $$\lim_{x \to 0} \chi = 0 \quad (z > 0)$$

Conditions (i) and (iv) are satisfied by

$$\chi(z', x') = \chi_0(z', x') + \Phi_1(z', x'),$$

where, by (6·10),
$$\chi_0(z', x') = \frac{\chi_s}{\Gamma(p)} \int_{z'^2/4x'}^{\infty} e^{-t} t^{p-1}\, dt$$

is the 'uncorrected' solution for the saturated strip and, from (8·1) with $\theta(\mu) = c'$,

$$\Phi_1(z', x') = \tfrac{1}{2} \chi_s (z'c')^p \int_0^{x'} \phi_1(\mu') \exp\left[-\frac{z'^2 + c'^2}{4(x' - \mu')} \right] I_p\left\{ \frac{z'c'}{2(x' - \mu')} \right\} \frac{d\mu'}{x' - \mu'}. \quad (9·1)$$

Conditions (ii) and (iv) are satisfied by

$$\chi(z, x) = \Phi(z, x) = \frac{\chi_s}{\sqrt{\pi}} \int_0^x \phi(\mu)\, (x - \mu)^{-\frac{1}{2}} \exp\left[-\frac{(z-c)^2}{4\sigma^2 (x - \mu)} \right] d\mu, \quad (9·2)$$

obtained from § 7 (11) by writing $p = \frac{1}{2}$ and $(z - c)/\sigma$ for z, or from the theory of the heat equation. It therefore remains to choose ϕ_1 and ϕ so as to satisfy the conditions of continuity at $z = c$. These are

$$\frac{1}{\sqrt{\pi}} \int_0^x \frac{\phi(\mu)\, d\mu}{(x - \mu)^{\frac{1}{2}}} = \frac{1}{\Gamma(p)} \int_{c'^2/4x'}^{\infty} e^{-t} t^{p-1}\, dt + \tfrac{1}{2} c'^{2p} \int_0^{x'} \phi_1(\mu')\, e^{-\omega'} I_p(\omega') \frac{d\mu'}{(x' - \mu')}, \quad (9·3$$

$$\frac{\sqrt{(A_c u_c)}}{B} \phi(x) = \frac{2^{1-2p} e^{-c'^2/4x'}}{\Gamma(p)\, x'^p} + \tfrac{1}{2} c'^2 \int_0^{x'} \phi_1(\mu')\, e^{-\omega'} \{ I_p(\omega') - I_{p-1}(\omega') \} \frac{d\mu'}{(x' - \mu')^2} - \tfrac{1}{2} \phi_1(x'), \quad (9·4)$$

where $\omega' = \frac{1}{2}c'^2/(x'-\mu')$. The equation (9·4) is obtained from (7·12), (8·5) and (8·7). The exact solution of these equations is likely to prove difficult, if not intractable but it is possible to obtain fairly simply a result for the orders of magnitude of ϕ_1 and ϕ when c' is large and $0 < p < \frac{1}{2}$. For if we use the asymptotic expansion of the terms on the right of (9·3), (9·4) and retain only the leading terms, we find

$$\int_0^x \frac{\phi(\mu)\,d\mu}{(x-\mu)^{\frac{1}{2}}} = \left[\frac{\sqrt{\pi}}{\Gamma(p)}\left(\frac{4x'}{c'^2}\right)^{1-p} e^{-\frac{1}{2}c'^2/x'}\right] + \tfrac{1}{2}c'^{2p-1}\int_0^{x'} \frac{\phi_1(\mu')\,d\mu'}{(x'-\mu')^{\frac{1}{2}}}, \tag{9·5}$$

$$\sqrt{\frac{(A_c u_c)}{B}}\phi(x) = \frac{2^{1-2p}e^{-\frac{1}{2}c'^2/x'}}{\Gamma(p)\,x'^p} + \left[\frac{1-2p}{4c'\sqrt{\pi}}\int_0^{x'} \frac{\phi_1(\mu')\,d\mu'}{(x'-\mu')^{\frac{1}{2}}}\right] - \tfrac{1}{2}\phi_1(x'), \tag{9·6}$$

the omitted terms being in each case of the order c'^{-2} compared with those retained. Now from (1·2) and (1·3) it is seen that $A_c u_c = a' u_1^{2-n} c$, proportional to c; also, from (1·5),

$$c' = 2(1-2p)\left(\frac{u_1^n}{a' x_0}\right)^{\frac{1}{2}} c^{(2-4p)^{-1}}. \tag{9·7}$$

Hence, using (1·8), we find $\sqrt{(A_c u_c)} = B\sqrt{(x_0)}\,c'^{1-2p}$.

It appears, therefore, that the terms enclosed in square brackets in (9·5), (9·6) can be neglected in comparison with the others when c' is large. The resulting equation (9·5) is then satisfied if

$$2\sqrt{x_0}\,\phi(x) = c'^{2p-1}\phi_1(x') \quad (x = x_0 x').$$

Substituting this value for $\phi(x)$ in (9·6) we obtain for the first approximation

$$\phi_1(x') = \frac{2^{1-2p}}{\Gamma(p)\sqrt{x'^p}}\,e^{-\frac{1}{2}c'^2/x'} > 0. \tag{9·8}$$

Now $$\operatorname*{Lim}_{z'\to 0}\left\{-z^{1-2p}\frac{\partial \Phi_1}{\partial z'}\right\} = \frac{-c'^{2p}\chi_s}{2^{2p}\Gamma(p)}\int_0^{x'} \frac{\phi_1(\mu')}{(x'-\mu')^{1+p}}\,e^{-\frac{1}{2}c'^2(x'-\mu')}\,d\mu'. \tag{9·9}$$

The correction to be applied to the rate of evaporation is therefore negative To obtain an estimate of its value we note that the expression obtained for $\phi_1(x')$ increases as x' increases from 0 to 1, c' being sufficiently large. The right-hand side of (9·9) is therefore numerically less than

$$\frac{\phi_1(x')\,c'^{2p}\chi_s}{2^{2p}\Gamma(p)}\int_0^{x'} \frac{e^{-\frac{1}{2}c'^2/(x'-\mu')}}{(x'-\mu')^{1+p}}\,d\mu' = \frac{\phi_1(x')\,\chi_s}{\Gamma(p)}\int_{\frac{1}{2}c'^2/x'}^{\infty} e^{-t}t^{p-1}\,dt \sim \frac{2^{3-4p}\chi_s}{\{\Gamma(p)\}^2}\,c'^{2p-2}x'^{1-2p}e^{-\frac{1}{2}c'^2/x'},$$

by (9·8) and the asymptotic expression for the integral. The resulting correction to the total rate of evaporation from the strip is therefore

$$\begin{aligned}
&-\frac{Bx_0 2^{3-4p}\chi_s}{\{\Gamma(p)\}^2}\,c'^{2p-2}\int_0^1 x'^{1-2p}e^{-\frac{1}{2}c'^2/x'}\,dx' \\
&= -\frac{Bx_0 2^{1-2p}\chi_s}{\{\Gamma(p)\}^2}\,c'^{2-2p}\int_{\frac{1}{2}c'^2}^{\infty} e^{-t}t^{2p-3}\,dt \\
&\sim -\frac{Bx_0 2^{4-4p}\chi_s}{\{\Gamma(p)\}^2}\,c'^{2p-4}e^{-\frac{1}{2}c'^2}.
\end{aligned}$$

The ratio of this to the total rate of evaporation calculated from χ_0 alone is

$$\{2^{3-2p}(1-p)/\Gamma(p)\}\, c'^{2p-4} e^{-\frac{1}{4}c'^2}. \tag{9·10}$$

In the case of the wind-tunnel experiments on evaporation discussed by Pasquill (1943) we take $c = z_1$, the height of the fully developed turbulent boundary layer, and $p = \frac{1}{6}$ corresponding to $n = \frac{1}{4}$. The value of the constant N occurring in the expression for a (§ 1) is 0·138 and we find from (9·7) with $c = z_1$, and cm.sec units,

$$c'^2 = 28\ 32 u_1^{\frac{1}{2}} z_1^{\frac{1}{2}} x_0^{-1}. \tag{9·11}$$

Hence with $u = 500$ cm./sec , $z = 1{\cdot}2$ cm., $x_0 = 20$ cm. we have $c'^2 = 8{\cdot}4$ Substituting this value in (9·10) we get for the correction the value 0·0017 This, of course, is the value of only the leading term in the correction, but it would seem unlikely that the full correction would amount to more than 1 or 2 % of the value obtained from χ_0 alone

In atmospheric experiments if an estimate can be obtained for the height z_1 up to which the power law (1·2) holds, then (9·11) may provide an estimate for the length x_0 downwind over which the effects of the correction may be disregarded.

10. The non-homogeneous equation

Expressions for an integral χ of (2·1) in terms of its boundary values and those of $z^{1-2p}\dfrac{\partial \chi}{\partial z}$ may also be derived from Green's formula (1·2), exactly as for the heat equation (Goursat 1923, § 545) As the non-homogeneous equation

$$L(\chi) = -f(z, x) \tag{10·1}$$

enters into some applications we shall obtain the results for it.

In figure 1 we replace the co-ordinates z, x by λ, μ and take AB, PQ as the characteristics $\mu = h$, $\mu = x$ respectively. Let R be an interior point (z, x) of PQ and R' a neighbouring point $(z, x+x_1)$, where $x_1 > 0$. In Green's formula, with λ, μ written for z, x, we take D as the region $ABPQA$, $U(\lambda, \mu)$ an integral of the equation (10·1) in λ, μ which is regular in D and V as $\chi_1(z, x+x_1; \lambda, \mu)$. Hence $M(V) = 0$, since χ_1 considered as a function of λ, μ satisfies the adjoint equation. Then with C the contour $APQBA$ we find

$$\int_{\overrightarrow{PQ}} \lambda^{1-2p} U(\lambda, x)\, \chi_1(z, x+x_1; \lambda, x)\, d\lambda = \iint_D f(\lambda, \mu)\, \chi_1(z, x+x_1, \lambda, \mu)\, d\lambda\, d\mu$$

$$+ \int_{\overrightarrow{PABQ}} \lambda^{1-2p} \left\{ U(\lambda, \mu)\, \chi_1 d\lambda + \left(\chi_1 \frac{\partial U}{\partial \lambda} - U \frac{\partial \chi_1}{\partial \lambda} \right) d\mu \right\}.$$

Now let $x_1 \to +0$. Then by § 4 the limit of the left-hand side is $U(z, x)$. The limits of the other terms follow by continuity and we obtain the fundamental formula:

$$U(z,x) = \iint_D f(\lambda,\mu)\,\chi_1(z,x;\,\lambda,\mu)\,d\lambda\,d\mu + \int_{AB} \lambda^{1-2p} U(\lambda,h)\,\chi_1(z,x;\,\lambda,h)\,d\lambda$$
$$+ \int_{AP+BQ} \lambda^{1-2p}\left\{U\chi_1 d\lambda + \left(\chi_1 \frac{\partial U}{\partial \lambda} - U\frac{\partial \chi_1}{\partial \lambda}\right)d\mu\right\}. \qquad (10{\cdot}2)$$

A similar formula, with χ_1 replaced by χ_2, is obtained from the second basic solution.

In the particular case where AE, BF are respectively $z = 0$ and $z = \infty$ the results simplify and yield direct solutions of the boundary value problems (i) and (ii) for the non-homogeneous equation (10·1). For $\operatorname{Lim}_{\lambda \to 0} \chi_1 = 0$ and on evaluating $\partial\chi_1/\partial\lambda$ from (5·5) by interchanging λ and z we find*

$$\operatorname{Lim}_{\lambda \to 0}\left\{\lambda^{1-2p}\frac{\partial \chi_1}{\partial \lambda}\right\} = \frac{(\frac{1}{2}z)^{2p}\,e^{-\frac{1}{4}z^2/(x-h)}}{\Gamma(p)\,(x-h)^{1+p}} = \{2^{-2p}/\Gamma(p)\}\,\chi_1(z,x;\,\mu).$$

Hence, in the notation of §§ 4, 7 we get from (10·2)

$$U(z,x) = \int_h^x \int_0^\infty f(\lambda,\mu)\,\chi_1(z,x;\,\lambda,\mu)\,d\lambda\,d\mu + F_1(z,x) + \Psi(z,x), \qquad (10{\cdot}3)$$

where F_1 is formed with $f(\lambda) = U(\lambda, h)$ and Ψ with $\psi(\mu) = U(0, \mu)$, the values of U on the boundaries $x = h$ and $z = 0$ respectively. Similarly we derive the solution of the second boundary value problem as

$$U(z,x) = \int_h^x \int_0^\infty f(\lambda,\mu)\,\chi_2(z,x;\,\lambda,\mu)\,d\lambda\,d\mu + F_2(z,x) + \Phi(z,x). \qquad (10{\cdot}4)$$

The formula (10·3) has been used to obtain an estimate of the error involved in neglecting the term $\partial\chi/\partial t$ in the non-steady equation

$$\frac{\partial}{\partial z}\left\{A_s \frac{\partial \chi}{\partial z}\right\} - u\frac{\partial \chi}{\partial x} = \frac{\partial \chi}{\partial t}, \qquad (10{\cdot}5)$$

where z, x are now actual space co-ordinates and t is the time. For if we regard (10·5) as a non-homogeneous equation it may be reduced to the form (10·1) where f involves $\partial\chi/\partial t$. The formula (10·3) then yields an integral equation for χ from which the order of magnitude of the error in the calculated drying time of a surface, made by putting $f = 0$, may be estimated. It is assumed, however, that a suitable expression for the slowly varying vapour concentration at the drying surface is known, and the formulation of such an expression is the chief difficulty met with in such applications of the theory.

$$\chi_1(z,x;\,\lambda,\mu) = \chi_1(\lambda,x;\,z,\mu).$$

I wish to express my thanks to Mr O. G. Sutton, B.Sc., and to Mr F. Pasquill, B.Sc., of the Meteorological Office, for their helpful discussion of matters arising during the preparation of this paper. Finally, acknowledgement is made to the Controller-General of Research and Development, Ministry of Supply, for permission to submit this paper for publication.

References

Bromwich 1908 *Infinite series.*
Giblett 1921 *Proc. Roy Soc.* A, **99**, 472
Goursat 1923 *Cours d'analyse Math.* **3**, ch. 29.
Pasquill 1943 *Proc. Roy. Soc* A, **182**, 75
Riemann-Weber 1901 *Partielle Diff.-Gleichungen*, **2**, § 38.
Sutton, O G. 1934 *Proc. Roy. Soc.* A, **146**, 701, see also Brunt, 1939, *Physical and Dynamical Meteorology*, chapter 12.
Taylor 1922 *Proc. Lond. Math. Soc.* **20**, 196.
Volterra & Pérès 1936 *Théorie générale des fonctionelles*, **1**, ch. 7
Watson 1922 *Theory of Bessel functions*, 13 31 (1).
Whittaker & Watson 1920 *Modern Analysis*, 17·7.

Evaporation from a plane, free-liquid surface into a turbulent air stream

By F. Pasquill, B.Sc

(*Communicated by D. Brunt, F.R.S.—Received 7 November* 1942)

In the past the experimental study of evaporation, and the interpretation of the results obtained, have shown in general only meagre reference to the aerodynamics of the problem. The present treatment, which is concerned with the evaporation from plane, free-liquid surfaces of relatively small dimensions into a tangential air stream, demonstrates the importance of the type of boundary layer flow The rates of evaporation under the influence of a turbulent boundary layer are then tested against a hydrodynamical theory due to O. G. Sutton.

Sutton's theory assumes that the turbulent transfer of any entity is determined by the momentum interchange coefficient, which is shown to involve the kinematic viscosity of the diffusing medium, and which leads to a functional form for evaporation which has been shown previously to be in good agreement with experimental data Developed into a computable form, and tested against the present experiments on the evaporation of bromobenzene and against experiments by Elias on the analogous problem of convective heat transfer, the theory is now shown to predict the absolute rate of turbulent transfer in a satisfactory manner.

An extension of the analysis to the relative rates of evaporation of various liquids, as determined in the present experiments and in recent experiments by Wade, shows that the theory specifies inadequately the variation of rate of evaporation with type of liquid. In the absence of a precise theoretical argument, an empirical generalization of Sutton's theory is set forth, in which the turbulent interchange coefficient is modified by the molecular diffusion coefficient appropriate to the entity undergoing transfer The range of physical characteristics covered in the present evaporation experiments, and in those performed by Wade, is sufficient to demonstrate the closer agreement provided by the generalized form of the theory. A more general test, against previous investigations for which the aerodynamic conditions can be estimated with reasonable confidence, shows that the absolute rate of evaporation may be predicted correctly in order of magnitude. In all cases considered the observed rates are in excess of the theoretical values, but it is significant that the discrepancy decreases as the experimental conditions conform more closely to the ideal conditions assumed in the theoretical treatment.

Introduction

1. In the numerous experimental researches on the evaporation from a plane, liquid surface, and in various attempts to collate the results of different workers, there occurs only meagre reference to the basic aerodynamics of the problem. By employing non-dimensional parameters, Powell (1942) has shown the results of many workers to be in encouraging agreement, but the only known example of interpretation on the basis of mathematical theory is due to Millar (1937), who employed the treatment of eddy transfer developed by Sverdrup (1936), and in order to obtain good agreement between theory and experiment Millar found it necessary to adjust the value of a constant in von Kármán's (1934) expression for skin friction

The present analysis of the problem includes considerable reference to the conditions of airflow over the evaporating surface, and the results are examined in the light of a theory of turbulent transfer due to O G Sutton (1934) The experimental work described is incomplete in that it is restricted to liquid surfaces of a few hundred square centimetres in extent, and to airflow which is typical of turbulent flow over smooth surfaces in wind tunnels and ducts, and the proposed generalization of Sutton's theory is largely dependent on general physical arguments. It is suggested, however, that the treatment constitutes an approach towards a comprehensive reduction of the problem, and also provides a means of calculating absolute magnitudes in evaporation under certain prescribed circumstances. Finally, in consideration of the broader implications of the theoretical treatment, opportunity is taken of analysing certain relevant data on heat transfer due to Elias (1929, 1930).

Experimental

2. In the experimental treatment of the problem the evaporating area takes the form of either a true free-liquid surface, in which case the liquid is contained in a pan, or of a thin sheet of absorbent material impregnated with the liquid. The pan technique suffers from the disadvantage of possible spillage in manipulation, except at the expense of incurring 'ridge' effects when the liquid is maintained at a safe level below the rim of the pan, although this difficulty may be overcome by the method adopted by Hine (1924), in which the evaporating pan is fed continuously with liquid, and arranged to overflow into a larger pan underneath. The impregnated absorbent surface cannot be regarded without restriction as representative of a free-liquid surface, but there appears to be no difficulty in providing an adequate practical approximation, and the risks of spillage and ridge effects are thereby largely obviated. From observations of the wet-bulb temperatures of thermometers covered with a free-water surface, and with linen, filter paper and gelatine impregnated with water, Powell (1942) has shown that the vapour pressure at the surface of moistened filter paper and similar materials does not differ appreciably from that at the surface of free water. In the present investigation with various

liquids, it was found that a Whatman no. 1 filter paper initially flooded with liquid gave rise to a sensibly constant rate of evaporation for a considerable portion of the drying period. This constant rate of evaporation is most naturally identified with the free-liquid rate of evaporation.

The method of estimation most frequently adopted involves a straightforward gravimetric or volumetric measurement of the loss due to evaporation. In Powell & Griffiths's experiments (1935), however, water was fed at a known rate to a linen surface and the excess water draining off was collected and measured, while in an alternative method employed by the same workers the rate of evaporation was derived from the difference between the electrical energy required to maintain the wet surface at a specified temperature, and that required when no evaporation was taking place. Wade (1942), in a more recent investigation with various organic liquids, adjusted a continuous feed of liquid so as to maintain the level of liquid flush with the rim of the container.

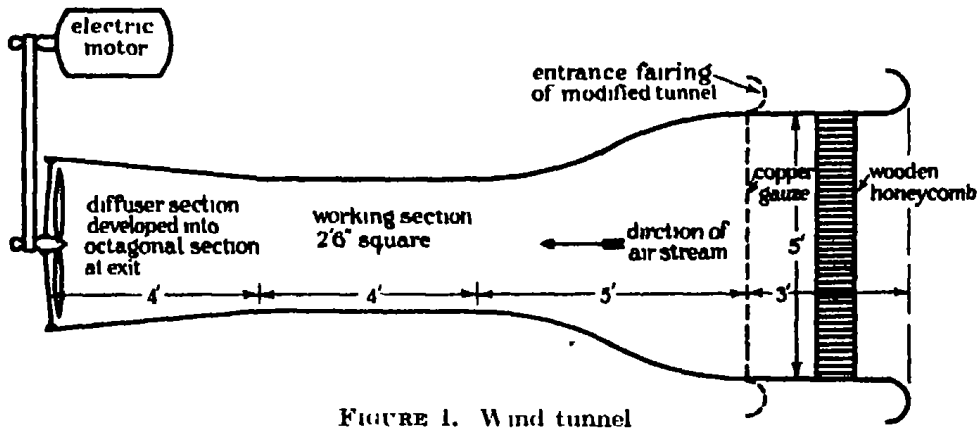


FIGURE 1. Wind tunnel

The aim in the present investigation was to employ the simplest apparatus and technique available, and the direct gravimetric method was applied to the case of a wet filter paper stretched on a glass plate Whatman no. 1 filter papers were employed, and these were maintained sufficiently moist to ensure that the rate of evaporation was appropriate to that from a free-liquid surface The periods of evaporation were adjusted to give a loss of between 2 and 3 g of liquid, and the loss was measured by an indicating balance which direct test had shown to be reliable to 0·03 g. In order to minimize the delays between weighing and exposure, the indicating balance was set up as closely as possible to the wind tunnel, and control measurements verified that the errors so introduced were within the limits of the experimental error in weighing.

The wind tunnel, in its original form, is shown diagrammatically in figure 1, and the various types of evaporation plates in figure 2. The section containing the wooden honeycomb and gauze was dispensed with after certain preliminary experiments, for reasons which are discussed later, and a new faired entrance was built on in the position indicated by the broken lines in figure 1.

3. While the preliminary experiments are not the main concern of the present analysis, it is proposed to give a brief discussion of the results obtained in view of the appearance of certain effects which have not been noted in the majority of previous investigations. These experiments employed the evaporation plate shown in figure 2A, and were concerned with the rate of evaporation of bromobenzene from an area 20 cm. across wind by 10 cm. downwind, as a function of wind speed and position on the evaporating plate. Relative rates of evaporation, corrected for temperature variations according to the principles discussed later, are shown in figure 3. The notable feature is that whereas the results of previous workers relate

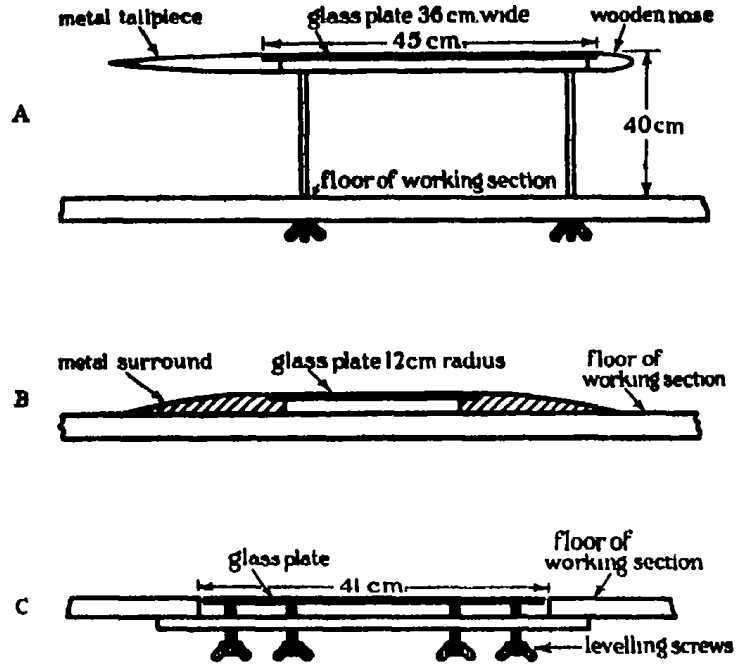


FIGURE 2. Evaporation plates

rate of evaporation and wind speed by a simple power law, the present results show a marked rise in rate of evaporation at a wind speed which appears to depend on the position of the evaporating strip on the glass plate. It was most natural to identify this effect with the transition from laminar to turbulent flow, and the existence of critical conditions of flow was indeed demonstrated by photographic observation of smoke flowing over the evaporating plate. Further verification may be sought in the investigation due to Dryden (1936*a*) on the transition from laminar to turbulent flow on an aerodynamically smooth flat plate exposed tangentially to air streams of different degrees of turbulence. For u'/u equal to 0·005, where u is the free-stream velocity and u' denotes the root mean square of the turbulent velocity component in the direction of the main stream, Dryden found that transition began at a distance x from the leading edge of the plate when the Reynolds number ux/ν was equal to 10^6. In the present investigation the value of u'/u was 0·0049, according to measurements made by Mr L. F. G. Simmons of the National Physical Laboratory,

but the relatively rough leading edge and filter-paper surface are conducive to a more rapid onset of transition, and on this basis the Reynolds number of 10^5 deduced from the results of figure 3 is in reasonable accord with the value given by Dryden.

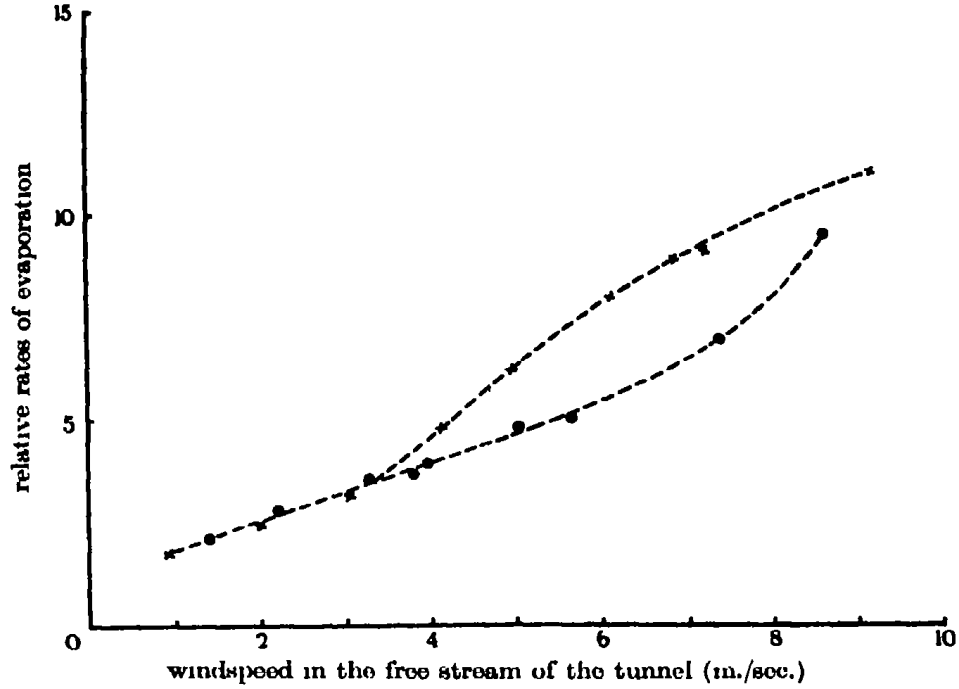


FIGURE 3 Relative rates of evaporation from a strip 20 cm. crosswind by 10 cm downwind on a free stream plate (see figure 2 A) Distance of downwind edge of strip from leading edge of evaporation plate—18 cm. (○) and 47 cm (×).

The above results demonstrate the necessity of directing attention to the aerodynamics of the experiment, and it is indeed surprising that in general similar effects have not been described hitherto. So far as can be ascertained the only previous reference is that due to Skljarenko & Baranajew (1936), who measured the rates of evaporation of several liquids from a surface of area 4 sq cm., and found that the relative rates of evaporation exhibited a sudden change with the transition from laminar to turbulent flow in the air stream It is, however, conceivable that most previous investigators have employed a boundary layer either laminar or turbulent, but never transitional

4. Since the main interest of the present investigation lay in the turbulent flow which occurs naturally in pipes and over plane surfaces, an attempt was made to impose this type of flow over the evaporating surface. Various methods were available, such as:

(i) the use of a longer evaporation plate,

(ii) the hastening of transition by the insertion of a bluff obstacle (e.g. a wire) in the laminar boundary layer,

(iii) the increase of the turbulence in the air stream,

(iv) the use of the boundary layer developed at the wall of the wind tunnel.

Method (i) was impracticable, and (ii) gave difficulties owing to the vigorous turbulence in the wake of the wire. The most satisfactory result followed from a combination of (iii) and (iv). A higher degree of turbulence was introduced in the free stream by merely removing the whole section containing the gauze and honeycomb, and the evaporating surface was made effectively part of the wind-tunnel wall by using either of the arrangements shown in figure 2 B and C. As there was no significant difference in the results obtained with these two types of evaporation plate, no further distinction will be made, and both types may be designated by the term 'wall plate', in contrast to the 'free-stream plate' shown in figure 2 A.

The complete investigation included the determination of the rate of evaporation of bromobenzene as a function of wind speed for three surfaces, namely,

circular, 12 cm. radius,
square, 19 by 19 cm.,
rectangular, 20 cm. across wind by 10 cm downwind,

followed by the determination of the rates of evaporation of bromobenzene, aniline, methyl salicylate and water from the circular area at a constant wind speed. With the exception of water, which was distilled in the laboratory as required, B D.H. reagents were employed in all cases. A summary of the vapour pressures adopted in the analysis of the results is given in table 1

TABLE 1. VAPOUR-PRESSURE DATA

	saturation vapour pressure in mm. of mercury at liquid temperatures in °F						
liquid	30	40	50	60	70	80	source of information
bromobenzene	0·73	1·10	1·62	2 32	3 31	4·65	extrapolation from *Intern. critical tables* (values for 30–60° C), and *Organic solvents* (O.U P. 1935) (value for 20° C)
aniline	0·071	0·115	0·180	0 284	0·434	0·663	extrapolation from values given by von Rechenberg for the range 20–50° C (*Einfache und fractionierte Destillation in Theorie und Praxis*, 1923)
methyl salicylate	0·014	0·023	0·037	0·061	0 095	0 147	Not published
water	4·18	6·29	9·22	13·3	18·8	26·3	*Smithsonian Physical Tables*

During evaporation experiments the room containing the wind tunnel was ventilated so as to avoid the building up of vapour concentrations of a magnitude sufficient to influence the rate of evaporation. Control observations of wind speed at the downstream end of the working section, and of air temperature in the working section, were carried out during each exposure of the evaporating surface, the former observation employing a sensitive cup anemometer of the type described by Sheppard (1940). In conjunction with experiments with water, wet- and dry-bulb

temperatures were also observed by means of an Assmann psychrometer mounted upstream of the evaporation plate. Subsequently the above control observations were related to the more appropriate factors of liquid-surface temperature and wind-velocity profile over the evaporation plate. An approach to the true surface temperature of the evaporating surface was obtained by using a 'surface thermometer' of the Negretti and Zambra type, which is essentially an ordinary mercury-in-glass thermometer with a thin flat metal disk attached to the thermometer bulb. The wind-velocity profile in the first few centimetres above the centre of the evaporating surface was measured by a small pitot-static tube of outside diameter 0·3 cm., used in conjunction with a Chattock-Fry gauge.

5. In view of the general lack of attention paid by previous workers to the question of velocity profile over the evaporating surface, it is desirable to discuss this point briefly with reference to the results of the present investigation

Employing the experimental resistance law for turbulent flow through tubes and assuming for the velocity profile the convenient form

$$u = u_{\max}(y/r)^q,$$

where r is the radius of the tube and y represents the distance from the wall, von Kármán (1921) and Prandtl (1925) deduced the value of $\frac{1}{7}$ for q This relation has been substantially verified by experiment, although it is now known that the index q decreases with Reynolds numbers above about 50,000. By analogy with flow through tubes, von Kármán (1921) and Prandtl (1927) assume the 'one-seventh' power law to apply to turbulent flow over a smooth flat plate, and although this relation has since been superseded by a more rigorous formula having theoretical basis (von Kármán 1934), the approximation is still regarded as adequate in many problems

The present observations of the velocity profile demonstrate a law of the type discussed above for the height range 0·4–1·2 cm. above the evaporating surface. No observations were made below 0·4 cm. owing to the possibility of serious interference between the pitot-static tube and the surface, while above 1·2 cm. there was evidently a transition region between the boundary layer profile and the constant free-stream velocity. Two distinct series of observations were carried out, separated by a period of more than a year, the appropriate values of q being 0·135 and 0·123, in reasonable agreement with the customary value of $\frac{1}{7}$.

Sutton's theory of evaporation in a turbulent atmosphere

6. Sutton's analysis of the problem of turbulent diffusion is based upon the adoption of an empirical expression for the correlation coefficient $R_\xi(z)$ between the vertical component of eddy velocity associated with a moving mass of fluid at time t, and the vertical component at time $t+\xi$, viz.

$$R_\xi(z) = \left(\frac{\lambda}{\lambda + \overline{w'^2}\xi}\right)^n, \tag{1}$$

where $\overline{w'^2}$ is the mean square eddying velocity in the vertical direction, λ is a parameter of dimensions cm.^{2} sec.$^{-1}$, and n is an indicator of the degree of turbulence present in the medium.

Making use of G. I. Taylor's work on 'Diffusion by continuous movements' (1922), Sutton then proceeded to obtain an explicit expression for the momentum interchange coefficient, and deduced the law of variation of wind with height

$$\frac{\bar{u}}{u_1} = \left(\frac{z}{z_1}\right)^{n/(2-n)}, \tag{2}$$

where $\bar{u}$ and $\bar{u}_1$ are the mean wind velocities at heights z and z_1 respectively.

Employing equation (2), the momentum interchange coefficient may be written

$$A(z)_{\text{momentum}} = \left|\frac{(\frac{1}{2}\pi k^2)^{1-n}(2-n)^{1-n}n^{1-n}}{(1-n)(2n-2)^{2-2n}}\right|\rho\bar{u}^{1-n}z^{1-n}\lambda^n, \tag{3}$$

where ρ is the air density and k is the so-called von Kármán's constant whose value according to Nikuradse (1932) is approximately 0·4.

Substituting (3) in the equation determining the shearing stress τ in the air in a smooth pipe, it was then shown that

$$\tau = \frac{0{\cdot}020}{z_1^{\frac{1}{4}}}\rho\lambda^{\frac{1}{4}}\bar{u}_1^{\frac{7}{4}},$$

as compared with the empirical expression

$$\tau = \frac{0{\cdot}023}{z_1^{\frac{1}{4}}}\rho\nu^{\frac{1}{4}}\bar{u}_1^{\frac{7}{4}},$$

a result which enabled Sutton to identify the parameter λ as the kinematic viscosity of the air.

Having established that the general development of the theory was in reasonable accord with fact, O. G. Sutton then turned to the two-dimensional problem of evaporation, generalizing the earlier work of Jeffreys (1918), who employed a constant interchange coefficient and a wind independent of height, by assuming that the transfer of mass in a turbulent atmosphere obeys the same laws, and is governed by the same interchange coefficient as the transfer of momentum.

7. In the steady case, the two-dimensional equation for diffusion of vapour from an evaporating surface level with the surface of the earth is

$$\bar{u}\frac{\partial V}{\partial x} = \frac{1}{\rho}\frac{\partial}{\partial z}\left\{A(z)\frac{\partial V}{\partial z}\right\}, \tag{4}$$

where x is measured downwind from the windward edge of the evaporating strip, z is measured vertically upwards, and V is the mass of vapour per unit mass of air

which is due to evaporation from the strip. Assuming the air density to be independent of height, equation (4) may be written

$$\bar{u}\frac{\partial \chi}{\partial x} = \frac{1}{\rho}\frac{\partial}{\partial z}\left\{A(z)\frac{\partial \chi}{\partial z}\right\}, \tag{5}$$

where χ is the mass of vapour per unit volume of air.

Subject to a boundary condition of constant vapour density at the evaporating surface, Sutton (1934, pp. 715–717) obtained the solution for equation (5) which, while not in a form suitable for computation, enabled the deduction of a functional form for the rate of evaporation $E(\bar{u}, x_0)$ from unit width of an infinite strip of length x_0 downwind, as follows

$$E(\bar{u}, x_0) = \int_0^\infty \bar{u}(\chi_{x=x_0})\,dz = K u_1^{(2-n)/(2+n)} x_0^{2/(2+n)}. \tag{6}$$

Neglecting lateral diffusion we have for a rectangular area of length x_0 downwind and width y_0 across wind

$$E(x_0, y_0) = K u_1^{(2-n)/(2+n)} x_0^{2/(2+n)} y_0, \tag{7}$$

and for a circular area of radius r

$$E(r) = K' \bar{u}_1^{(2-n)/(2+n)} r^{(4+n)/(2+n)}, \tag{8}$$

the quantities K and K' being independent of $\bar{u}_1$, x_0 and r.

It has been shown* that the original solution may be expressed in the form

$$\chi = \chi_0\left[1 - \frac{1}{\pi}\sin\frac{2\pi}{2+n}\,\Gamma\left(\frac{2}{2+n}\right)\Gamma\left\{\frac{\bar{u}_1^n z^{(2+n)/(2-n)}}{\left(\frac{2+n}{2-n}\right)^2 a z_1^{n/(2-n)} x}, \quad \frac{n}{2+n}\right\}\right], \tag{9}$$

where

$$a = \left|\frac{(\frac{1}{2}\pi k^2)^{1-n}(2-n)^{1-n}n^{1-n}}{(1-n)(2n-2)^{2-2n}}\right| \lambda^n z_1^{(n^2-n)/(2-n)}, \tag{10}$$

$$\chi_0 = \lim_{z\to 0}\chi(x, z) \quad (0 < x < x_0),$$

$$\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}\,dx \quad \text{and} \quad \Gamma(\theta, p) = \int_0^\theta x^{p-1}e^{-x}\,dx,$$

* O. G. Sutton & K. L. Calder (unpublished). The derivation depends upon the following identity in Bessel functions, communicated privately by Dr F. J. W. Whipple. The identity, deduced from Weber's exponential integral and its generalizations (Watson, 1922, p. 394, eqs. (3) and (4)), is

$$\int_0^\infty K_\nu(\theta)\exp(-r^2\theta^2)\,\theta^{1-\nu}\,d\theta = \frac{\Gamma(1-\nu)}{2}(2r^2)^{\nu-1}\exp\left(\frac{1}{4r^2}\right)\int_{1/(4r^2)}^\infty e^{-\theta}\theta^{\nu-1}\,d\theta,$$

where $K_\nu(\theta)$ is the modified Bessel function of the second kind with purely imaginary argument (Watson 1922, p. 78).

and in the expressions (6), (7) and (8) for E,

$$K = \chi_0 \left(\frac{2+n}{2-n}\right)^{(2-n)(2+n)} \left(\frac{2+n}{2\pi}\right) \sin\frac{2\pi}{2+n}\, \Gamma\left(\frac{2}{2+n}\right) a^{2/(2+n)} z_1^{-n^2/(4-n^2)}, \tag{11}$$

$$K' = \frac{2^{2+n} K \sqrt{\pi}\, \Gamma\left(\frac{3+n}{2+n}\right)}{\Gamma\left\{\frac{8+3n}{2(2+n)}\right\}}. \tag{12}$$

The rate of evaporation may now be calculated from equations (7), (8), (10), (11) and (12), providing we have numerical values for n, $\bar{u}_1$, z_1, which follow from a knowledge of the wind-velocity profile, and for χ_0 and λ It is reasonable to identify χ_0 with the saturation vapour density at the temperature of the liquid surface, and if as a working approximation the perfect gas law be assumed to hold, then with p_s as the saturation vapour pressure, M the molecular weight of the liquid, R the universal gas constant and T the absolute temperature, we may evaluate χ_0 from the expression

$$\chi_0 = p_s M/RT. \tag{13}$$

According to Sutton the parameter λ is the kinematic viscosity of the air, so that numerical evaluation of the rate of evaporation is now possible. From equations (10), (11) and (12),

$$K \text{ (or } K') \propto \chi_0 \lambda^{2n/(2+n)} k^{4(1-n)/(2+n)} z_1^{-n/(2+n)},$$

and table 2 gives values of K and K' computed for $\chi_0 = 1$ g /c c., $\lambda = \nu = 0{\cdot}147$ cm.^{2} sec.$^{-1}$, $k = 0{\cdot}4$, $z_1 = 1$ cm. and n ranging from 0·2 to 0·3.

TABLE 2. VALUES OF K AND K' FOR $\chi_0 = 1$ G./C C., $\lambda = \nu = 0{\cdot}147$ CM.^{2} SEC.$^{-1}$, $k = 0{\cdot}4$, $z_1 = 1$ CM.

n	K	K'	n	K	K'
0·20	0·0094	0·0283	0·26	0·0202	0·0601
0·21	0 0108	0 0325	0·27	0 0226	0 0674
0·22	0·0124	0·0371	0 28	0 0253	0·0751
0·23	0·0141	0·0423	0 29	0·0282	0 0835
0 24	0 0160	0·0477	0·30	0·0313	0·0926
0 25	0·0180	0 0537			

Before proceeding, let us consider the principal difficulties arising in a comparison of the present hydrodynamical theory with experiment. The theory assumes a wind speed constant in the horizontal direction, and a power-law variation in the vertical. In many practical cases of airflow over a plane surface the momentum boundary layer is in an early stage of development. Accordingly, the thickness of the boundary layer, which may be of the order of only a few centimetres, and the wind velocity at a given distance from the surface, vary with horizontal position. Moreover, the characteristic turbulent velocity profile is limited to the region between the thin laminar sublayer at the boundary, and the region of transition from the true

boundary layer to the more or less constant free-stream velocity. Again, the application of the theory to evaporating strips of finite width presupposes the absence of lateral diffusion, but this is probably the least important of the discrepancies, since tests carried out during the present investigation indicate that for the areas employed the effects introduced by lateral diffusion amount to 3 or 4 % of the total rate of evaporation. On the above terms the theoretical treatment is clearly a considerable idealization of the general practical case, and the closeness of the approximation provided can only be assessed by the actual comparison with experiment.

Comparison of experiment with theory

8. The results of the main series of experiments with bromobenzene have been related to p_s, the saturation vapour pressure at the temperature of the liquid surface, and $u_{1\,cm}$, the wind velocity at a height of 1 cm. above the centre of the evaporating strip. Comparison of theory and experiment is facilitated by reducing all results to a common unit, namely, the value calculated from the theory for $u_{1\,cm}$ equal to 5 m./sec The appropriate values of n are obtained by equating the exponent in equation (2) to the observed value of the index in the power-law distribution of velocity with height. Then from equations (7), (8), (10), (11), (12) and (13), for $T = 290°$ A, $M = 157$, $R = 83{\cdot}15 \times 10^6$, $k = 0{\cdot}4$, $\lambda = \nu = 0{\cdot}147$, $z_1 = 1$ cm., $u_1 = 500$ cm./sec , we have

(*a*) circular area 12 cm. radius, $n = 0{\cdot}238$,
$E/p_s = 0{\cdot}357$ g./min /mm. of mercury;

(*b*) square 19×19 cm., $n = 0{\cdot}238$,
$E/p_s = 0{\cdot}286$ g./min /mm. of mercury,

(*c*) rectangle 20 cm. across wind, by 10 cm., $n = 0{\cdot}219$,
$E/p_s = 0{\cdot}149$ g./min./mm. of mercury.

Smoothed experimental results, reduced on the above basis, are reproduced in table 3, and the individual observations are shown graphically in figure 4, from which it is seen that there is excellent agreement between theory and experiment.

Table 3. Comparison of present observations with theory

$u_{1\,cm}$ in m./sec. ..	1	2	3	4	5	6	7	8	9
rate of evaporation on Sutton's theory	0·28	0·48	0 67	0·84	1·00	1·16	1·31	1·45	1·59
rate of evaporation observed	0·25	0·43	0·60	0·76	0 92	1 06	1·20	1·34	1·48

The comparison of experiment with theory may now be completed by considering merely the relative rates of evaporation of the various liquids, since from equations (6), (10), (11) and (13) the rate of evaporation E of a liquid should be proportional to Mp_s/T, other factors remaining constant. In the general case p_s should be replaced

by $(p_s - p_a)$, where p_a is the partial pressure of vapour in the air upwind of the evaporating surface. At this stage the results obtained in the present investigation may be supplemented by those obtained recently by Wade (1942), who employed a wind channel with a working section 12 cm. wide and 6 cm. high, and an evaporating pan arranged so that the liquid surface was flush with the floor of the working section. Taking Schiller's (1923) value of approximately 2000 for the Reynolds number appropriate to the transition to turbulent flow in a smooth rectangular pipe, it would appear reasonable to assume turbulent flow in Wade's experiments over the working range of wind speeds 1–4 m./sec.

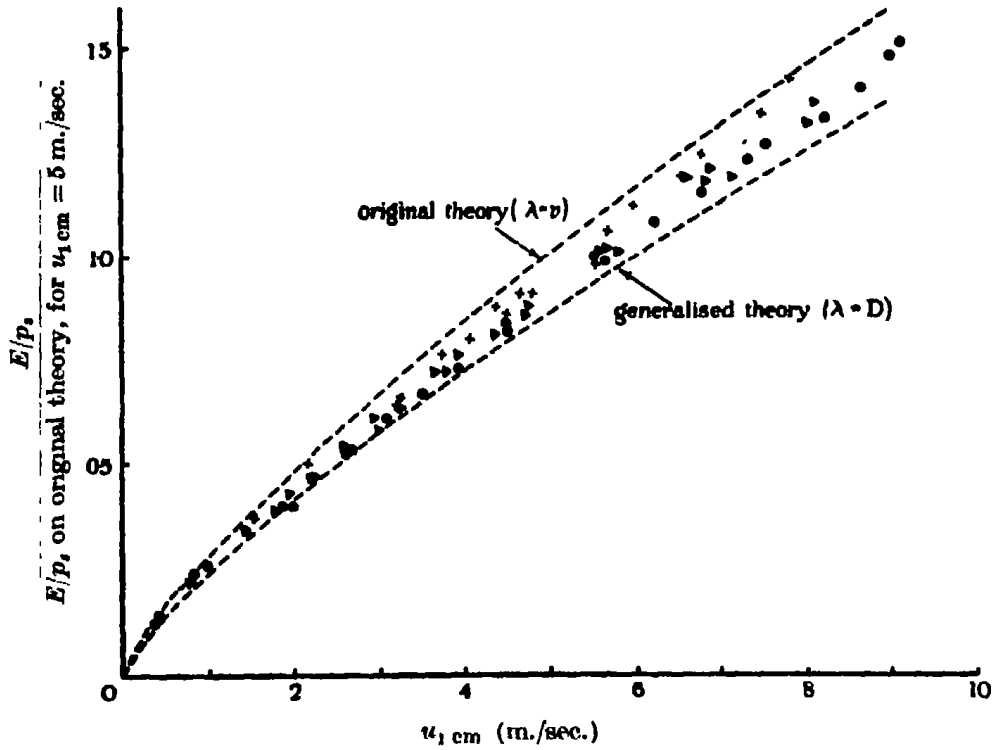


FIGURE 4. Experimental and theoretical values for the rate of evaporation of bromobenzene into a turbulent air stream ⊙ = observed values for surface 24 cm. diameter; △ = observed values for surface 19 cm. square; × = observed values for surface 20 cm. cross-wind by 10 cm. downwind.

The complete analysis, in which the two sets of results are reduced to a common basis by taking the case of water as standard, is shown in table 4. The normal variations of working temperature are not significant in the present case, and average values appropriate to the present investigation (290° A) and to Wade's experiments (310° A) are adopted.

The variations exhibited by the parameter $ET/[M(p_s - p_a)]$ denote a serious discrepancy in the implications of the theoretical treatment when applied to liquids with a wide range of physical characteristics. It is to be noted that if water be omitted from the above considerations the range of variation of the parameter is

considerably reduced, and it may be for this reason that previous workers, notably Hofmann (1932) and Hine (1924), conclude that rate of evaporation is proportional to the product $M(p_s - p_a)$. Wade summarizes his results conveniently by introducing the molecular weight term raised to a power of magnitude 0·71, a form of expression which cannot be regarded as possessing physical significance, and in an attempt to interpret the experimental facts in a more rigorous manner a generalization of Sutton's treatment is now set forth.

TABLE 4. RATES OF EVAPORATION OF VARIOUS LIQUIDS RELATIVE TO WATER

liquid	$ET/[(p_s-p_a)]$	M	$ET/[M(p_s-p_a)]$
water (A and W)	1·00	18	1·00
acetone (W)	2·52	58	0·78
benzene (W)	3·54	78	0·82
ethyl acetate (W)	3·10	88	0·63
toluene (W)	3·44	92	0·67
aniline (A)	3·82	93	0·74
trichlorethylene (W)	4·18	131	0·57
methyl salicylate (A)	5·09	152	0·60
carbon tetrachloride (W)	5·00	154	0·58
bromobenzene (A)	6·72	157	0·77

W, Wade's experiments, A, present experiments.

A GENERALIZATION OF SUTTON'S THEORY

9. The basic physical principle of O G. Sutton's treatment is that in a turbulent medium, an eddy, which is considered as a mass of fluid, leaves its surroundings under the action of some external disturbing force, and, moving as a discrete body, blends continuously with the surrounding medium. Consider now an atmosphere possessing some transferable conservative property such as vapour, heat content or momentum. As the surroundings of the eddy change there will be a continuous sharing of the transferable property, and a tendency to reduce the excess or defect possessed by the eddy, until the eddy becomes indistinguishable from its surroundings. At this stage 'mixing' is said to be complete.

Diffusion of heat, matter and momentum is therefore regarded as controlled primarily by the movement of eddies or masses of fluid, but since the sharing of these entities between an eddy and its surroundings must depend ultimately on molecular action, then it is inconceivable that the constants of molecular diffusion should not enter the general expressions for turbulent transfer. This principle is immediately verified by O. G. Sutton's analysis of shearing stress in a smooth pipe, on which basis the parameter λ appearing in equation (1) is identified with the kinematic viscosity of the air, but so far it has proved impossible to arrive at a precise theoretical definition of the form in which the molecular constants appear in the equations of turbulent transfer in general. We resort therefore to an empirical argument as follows.

10. On the basis of Sutton's analysis of the shearing stress in the air in a smooth pipe, the expression for the momentum interchange coefficient (equation (3)) may be re-written as

$$A(z)_{\text{momentum}} = \left| \frac{(\frac{1}{2}\pi k^2)^{1-n}(2-n)^{1-n}n^{1-n}}{(1-n)(2n-2)^{2-2n}} \right| \rho \bar{u}^{1-n} z^{1-n} \nu^n$$

$$= f(n)\rho\bar{u}z\left(\frac{\nu}{\bar{u}z}\right)^n, \tag{14}$$

where the quantity $\bar{u}z/\nu$ is the Reynolds number, which occurs naturally in all problems of turbulent motion. It may be noted at this stage that the above functional form of $A(z)$ is deducible from dimensional considerations, so that the main interest lies in the explicit function $f(n)$.

In the problem of convective heat transfer (Dryden 1936*b*) dimensional considerations suggest a large number of dimensionless combinations of the fundamental variables. For forced convection only, and assuming the effect of viscosity to be negligible, Dryden shows that the controlling factor is the Péclet number, which, in the notation appropriate to the Reynolds number above, may be written $\bar{u}z/\kappa$, κ being the thermal diffusivity. By analogy, for the case of vapour transfer, we may define the controlling factor as uz/D, which may be termed the diffusion number, where D is the vapour diffusivity.

The above considerations suggest that the interchange coefficient for heat should be

$$A(z)_{\text{heat}} = f(n)\rho\bar{u}z\left(\frac{\kappa}{\bar{u}z}\right)^n, \tag{15}$$

and for vapour we should expect

$$A(z)_{\text{vapour}} = f(n)\rho\bar{u}z\left(\frac{D}{\bar{u}z}\right)^n. \tag{16}$$

In the foregoing expressions for the interchange coefficient viscosity effects are implicitly involved in the sense that the quantity n is determined by the wind-velocity profile, but it may be argued that the kinematic viscosity should appear in more direct association with the thermal and vapour diffusivities, on the grounds that viscosity must influence the eddy transfer of any entity. However, in the absence of a developing boundary layer and of pronounced natural convection, conditions which are assumed in the theoretical problem and approximated to in the practical cases considered, it is considered unlikely that viscosity will be of more than secondary importance. From this standpoint equations (15) and (16) may be regarded as close approximations to more general expressions in which κ and D are dominant terms. In any case, the ultimate justification for the generalization must depend on the agreement between theory and experiment.

Further analysis of the experimental data

11. The only modification imposed on O. G. Sutton's solution for the evaporation case is the replacement of ν by D, and from equations (6), (10) and (11) it is seen that

the absolute rate of evaporation is modified in the ratio $(D/\nu)^{2n/(2+n)}$. Under the conditions of the main series of experiments with bromobenzene, $T = 290°$ A, $\nu = 0{\cdot}147$ cm.^{2} sec.$^{-1}$, D (vapour to air) $= 0{\cdot}068$ cm.^{2} sec.$^{-1}$, and n (average) $= 0{\cdot}23$. Values appropriate to the generalized theory may be obtained from the values on the original theory by multiplying the latter by $(D/\nu)^{2n/(2+n)}$, which in the present case is equal to 0·86. An inspection of table 2 shows that no significant change in the agreement with observed values would follow from an application of this factor, and as a demonstration of this point the curve appropriate to the generalized theory is included in figure 4.

From equations (6), (10), (11) and (13) we now see that the rate of evaporation of a liquid should be proportional to $[M(p_s-p_a)\,D^{2n/(2+n)}]/T$. In reducing Wade's results on this basis n is assumed to have the nominal value of 0·25, which is appropriate to the customary 'one-seventh' power-law velocity profile, and since the variations from this nominal value observed in the present investigation are not of great significance in the term $2n/(2+n)$, the value of 0·25 is adopted throughout the following analysis. In table 5 are given the previous values of $ET/[M(p_s-p_a)]$, with values of D appropriate to the average temperature of 310° A in Wade's experiments, and 290° A in the present investigation, and of the new parameter $ET/[M(p_s-p_a)\,D^{2n/(2+n)}]$. The results cover a range of 18–157 in molecular weight, 0·05–0·28 cm^2 sec.$^{-1}$ in vapour diffusion coefficient, and 0 04–400 mm. of mercury in vapour pressure. It is seen that a much greater degree of constancy is exhibited by the new parameter than by the parameter derived from Sutton's original solution, demonstrating clearly the superiority of the generalized form of the theory.

TABLE 5. RATES OF EVAPORATION OF VARIOUS LIQUIDS RELATIVE TO WATER

liquid	$ET/[M(p_s-p_a)]$	D	$ET/[M(p_s-p_a)\,D^{2n/(2+n)}]$ ($n = 0{\cdot}25$)
water (A and W)	1·00	0·275 (W) 0 244 (A)	1·00
acteone (W)	0·78	0·094*	0 99
benzene (W)	0·82	0·099	1 02
ethyl acetate (W)	0·63	0 093	0·80
toluene (W)	0·67	0·092	0·86
aniline (A)	0·74	0 069	0·98
trichlorethylene (W)	0·57	0·079*	0 75
methyl salicylate (A)	0·60	0·054*	0·84
carbon tetrachloride (W)	0 58	0·071*	0·79
bromobenzene (A)	0·77	0·068*	1·02
furfural (A)†	0·77	0 076	1·00

W, Wade's experiments; A, present experiments.

† Results obtained since submitting MS

The values of D were obtained from International Critical Tables or calculated (*) from the empirical formula due to Gilliland (1934), except in the case of furfural for which the value of D was measured recently (unpublished), and

correspond in all cases to a total pressure of 1 atmosphere. In neglecting the partial pressure of vapour the maximum error involved in the term $D^{2n/(2+n)}$ is 5 %, which, in view of the otherwise approximate nature of the calculated values of D, is considered unimportant in the present analysis.

12. On the basis of the above analysis one may expect to predict with an error of less than 20 % the evaporation of liquids of accurately known characteristics, in the case of small plane surfaces exposed to a tangential air stream typical of turbulent flow in wind tunnels and ducts. Assuming the 'one-seventh' power law of velocity distribution in the vertical, knowledge is then required of the wind speed u_1 at a height z_1 consistent with the latter profile. In many cases difficulty may be anticipated in defining u_1 and z_1 accurately, but it is unlikely that the errors involved will preclude an estimate of order of magnitude, which alone may be of considerable value in many practical problems. With this end in view an examination has been made of numerous experimental researches, and four cases have been found in which it seems reasonable to assume turbulent flow, and in which a fair estimate may be made of the quantities u_1 and z_1. The cases in point are those of Hine (1924), Hinchley & Himus (1924), Millar (1937) and Wade (1942), who employed without exception evaporation pans of the 'wall' type Details which are relevant to a comparison of observed and theoretical values are given in table 6.

Except in the case of Hine's investigation it is necessary to estimate the appropriate values of z_1 from the meagre data available. However, since z_1 appears in the form $z_1^{-n/(2+n)}$, with n equal to 0·25, a 500 % error in z_1 will introduce an error of only 20 % in the rate of evaporation, and such a magnitude of error in the estimation of z_1 is most unlikely in the cases under consideration. Hine and Wade both employed a selection of liquids, but the present comparison is restricted to those liquids for which the accuracy and number of observations was greatest, and for which accurate values of the vapour-diffusion coefficients are available Adopting the value of 0·25 for n, consistent with the one-seventh power law of velocity distribution, absolute rates of evaporation have been calculated as previously, taking λ equal to the vapour-diffusion coefficient D, and a comparison between observed and computed values is shown in table 7 and figure 5. The results appropriate to the investigation of Hinchley & Himus were actually extracted from a reanalysis of the data carried out by Himus (1929). In all cases smoothed results are employed, and to complete the comparison the present series of observations with bromobenzene (figure 4) is included.

The most striking feature of the comparison shown in table 7 and figure 5 is that all observed values considered are in excess of the theoretical value, and this is consistent with the reasonable supposition that in the case of evaporating pans the main sources of error will be the existence of ridges at the edges of the evaporating surface, and of ripples on the surface of the liquid. In general, both effects may be expected to result in an increased rate of evaporation, the influence of ridges being demonstrated clearly in Powell & Griffiths's investigation (1935), and it is to be anticipated that observed rates of evaporation will be in excess of the predicted values to a degree dependent on the departure from the ideal conditions of flow

TABLE 6. DETAILS OF THE EXPERIMENTAL INVESTIGATIONS TO BE COMPARED WITH THEORY

	investigators			
details	Hine	Hinchley & Himus	Millar	Wade
size of evaporating surfaces	59 cm. diam.	3 by 12 in. 6 by 12 in. 12 by 12 in. 6 by 18 in.	50·8 cm. downwind by 25·4 cm. across wind	8·9 cm. sq.
dimensions of working section of wind tunnel or duct	6 ft. sq.	1 ft. 6 in. wide by 9 in. high	38 cm. wide by 13 cm. high	12 cm. wide by 6 cm. high
length of wind tunnel or duct upwind of evaporating surface	280 ft.	5 ft.	not given	approximately 100 cm.
site of measurement of wind speed (adopted as u_1)	immediately downwind of evaporating surface; axis of vane anemometer at height of 2 in.	at centre of duct exit—size of anemometer not specified	air meter 10 cm. in diameter standing on floor of tunnel	orifice meter at exit; results expressed in terms of maximum velocity in tunnel from hot-wire anemometer exploration
range of wind speeds	0·5–3 m./sec.	1–6 m./sec.	1·5–4·5 m./sec.	1–4 m./sec.
value of z_1 adopted	5 cm.	4 cm.	1 cm.	1 cm.
liquids considered in present comparison	toluene	water	water	(*a*) ethyl acetate, (*b*) toluene
average temperature of liquid surface	287° A	313° A	not given, assumed 288° A	310° A
vapour diffusion coefficient (D), cm.^{2} sec.$^{-1}$, from *International critical tables*	0·078	0·28	0·24	(*a*) 0·093, (*b*) 0·092

presupposed in the theoretical treatment. It is accordingly of considerable significance that the agreement with the theoretical result is much better in the cases where there appears to be little chance of such spurious influences, that is, in the experiments performed by Wade and the present writer. Hinchley & Himus refer to the existence of a ridge at the edge of the pan, and to the occurrence of ripples; Hine states that ripples occurred over the whole working range of wind speed, and in addition the liquid surface was actually flush with the surface of a steel plate of thickness $\frac{1}{4}$ in. mounted on the floor of the duct and extending only some 10 cm.

upstream of the evaporating surface. Furthermore, in Hine's investigation the evaporating pan became warped, and a steel plate was placed on the edge of the pan to keep it level. It would appear, therefore, that the present theory gives a very satisfactory measure of agreement in the absence of incalculable effects such as those discussed, and that even when such effects are present, as they undoubtedly will be in many practical cases, the predicted rates of evaporation, although underestimates, are of the correct order of magnitude.

TABLE 7. COMPARISON OF VARIOUS OBSERVATIONS WITH GENERALIZED THEORY—THE THEORETICAL RATE OF EVAPORATION FOR A WIND VELOCITY OF 3 M./SEC. BEING TAKEN AS UNITY

wind speed in m./sec. / source	0·5	1·0	1·5	2·0	2·5	3·0	3·5	4·0	4·5	5·0	5·5	6·0
Hine	0·52	0·87	1·20	1·51	1·80	2·07	—	—	—	—	—	—
Hinchley & Himus	—	1·25	1·50	1·71	1·93	2·13	2·32	2·51	2·70	2·89	3·08	3·27
Millar	—	—	0·81	1·02	1·22	1·40	1·58	1·75	1·92	—	—	—
Wade	—	0·50	0·71	0·90	1·09	1·27	1·45	1·61	1·79	—	—	—
present experiments	—	0·43	0·59	0·75	0·90	1·04	1·18	1·32	1·46	1·60	1·72	1·84
theory (generalized)	0·25	0·42	0·58	0·73	0·87	1·0	1·13	1·25	1·37	1·49	1·60	1·72

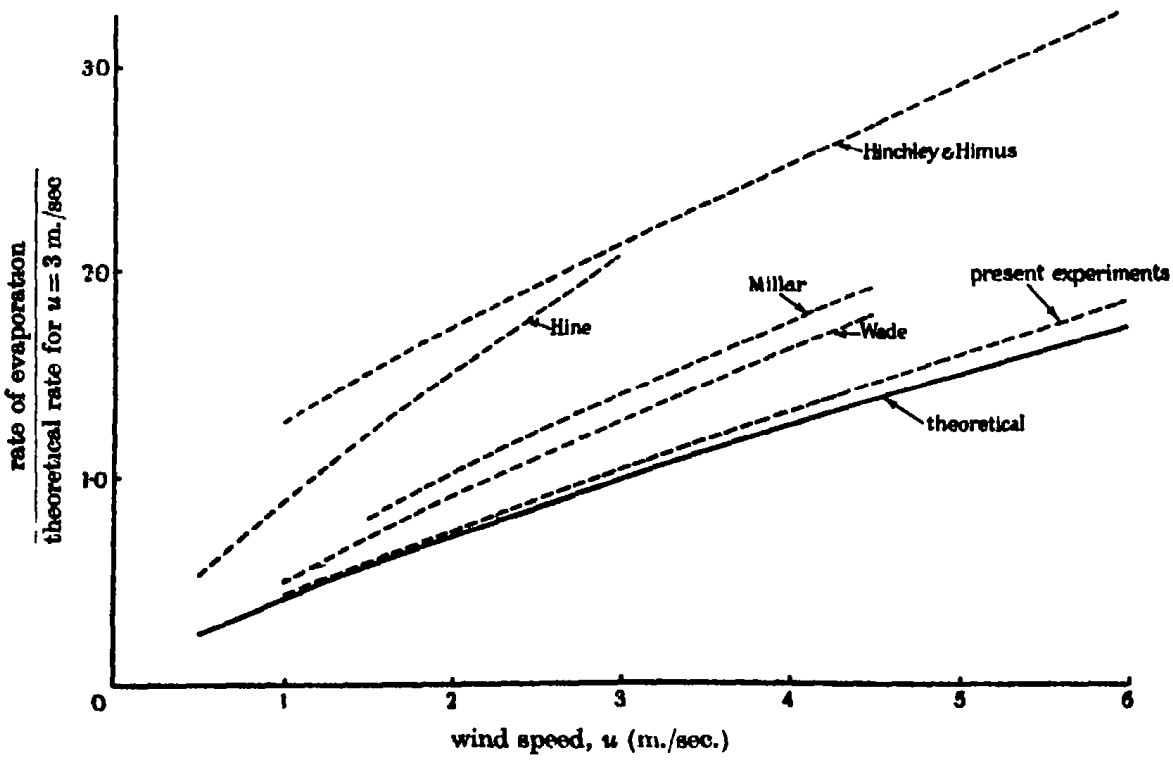


FIGURE 5. Rate of evaporation into a turbulent air stream; comparison of experimental and theoretical values.

Heat transfer from a flat plate by forced convection

13. The foregoing theoretical analysis may be applied to the turbulent diffusion of heat in the boundary layer over a smooth surface, in which case the concentration χ occurring in the evaporation problem is replaced by the appropriate function of temperature in the air stream. Equations (6)–(12) then prescribe the transfer of heat by forced convection when χ_0 is replaced by $\rho C_p t$, where C_p is the specific heat of air at constant pressure, and t is the difference in temperature between the plate surface and the free air.

On the original theory the parameter λ is replaced by the kinematic viscosity ν, while on the generalized version we write for λ the thermal diffusivity κ of the air. Owing to the closeness of the numerical values of ν and κ, it is clear that experimental results are unlikely to be of sufficient accuracy to differentiate between the two forms, and at the most we may only anticipate a test of the theory in general.

The rate of transfer of heat from a heated plate by forced convection has been studied experimentally by various workers, and the investigation most appropriate to the present analysis is that due to Elias (1929, 1930). Elias exposed a smooth copper plate, electrically heated and maintained at a constant temperature, to a tangential air stream in a wind tunnel, and made a detailed exploration of the temperature and wind velocity profiles over the plate. The temperature and wind velocity measurements were made with thermocouples and a fine glass pitot tube, the plate being mounted in a vertical plane, presumably in order to avoid the occurrence of natural convection in combination with the forced convection. It is noteworthy that in order to approximate closely to the two-dimensional case, the observations were made on the axis of the plate which was 25 cm. wide, and, moreover, it was observed that the velocity profile was not influenced significantly by the heating of the plate.

Elias found that the velocity and temperature profiles in turbulent flow were practically identical, and by computing the heat transfer, Q, from the observed distributions of temperature and velocity, showed that

$$Q \propto u^{0\cdot8} x_0^{0\cdot89},$$

a result which is in excellent agreement with equation (6) when $n = 0{\cdot}25$, and which has been quoted previously (Brunt 1939; Lettau 1939). It is noteworthy that in an analysis of the experimental results of various workers, Fishenden & Saunders (1932) conclude that the heat loss is proportional to a power of u not far from 0·8.

From an examination of the experimental wind velocity profile obtained by Elias for a free-stream velocity of 35 m./sec., it is evident that the 'one-seventh' power law is satisfied up to a distance of at least 2 mm. from the plate over the whole length of the heated portion. The 2 mm. value corresponds to the upwind edge of the heated zone, and proceeding downwind the profile is extended to greater distances from the plate. Thus in applying equations (6), (10) and (11) we take $n = 0{\cdot}25$ and

$z_1 = 0{\cdot}2$ cm. Furthermore, the observations show that u_1 decreases from 34 m./sec. at the upwind edge to 26·5 m./sec. at the downwind edge, with an average value of 29·5 m./sec. over the whole heated zone.

Taking $\rho = 1130 \times 10^{-6}$ g./c.c., $C_p = 0{\cdot}237$, $\nu = 0{\cdot}17$ cm.^{2} sec.$^{-1}$, $\kappa = 0{\cdot}24$ cm.^{2} sec.$^{-1}$, for a temperature of 40° C, which approximates closely to the temperature of the plate, the theoretical values of Q, the heat loss from unit width of an area 40 cm. downwind for $t = 1°$ C and $u_1 = 29{\cdot}5$ m. sec., are

$8{\cdot}0 \times 10^{-3}$ cal./sec.—on the original theory

and $8{\cdot}6 \times 10^{-3}$ cal./sec.—on the generalized theory

as compared with the observed value of

$11{\cdot}7 \times 10^{-3}$ cal./sec.,

an agreement which provides a substantial verification of the general theoretical approach.

In the above calculation a more rigorous procedure would employ a weighted mean wind velocity, and would result in a higher theoretical value of Q However, owing to the closeness to unity of the exponent in the power-law relationship between Q and x_0, the difference between the weighted and arithmetic means is insufficient to modify the agreement in a significant sense.

The writer is indebted to Professor D. Brunt, F.R.S., Professor S. Sugden, F.R.S., Mr O. G. Sutton, B.Sc., and Mr H. Garnett, M.Sc., for valuable encouragement and criticism during the preparation of this paper. Thanks are also due to Mr S. G. Crawford, B.Sc., and Mr S. C. Mason, of the Meteorological Office, for assistance in the experimental work, to Mr J. H. Simpson, M.Sc., of the Meteorological Office, for checking the analysis and computations, and to Dr S. H Wade for permission to employ experimental results which have not yet been published in full Finally, acknowledgement is made to the Controller-General of Research and Development, Ministry of Supply, for permission to submit this paper for publication.

References

Brunt 1939 *Physical and dynamical meteorology*, 2nd ed. p. 275.
Dryden 1936*a* *Rep. Nat. Adv. Comm. Aero., Wash.*, no. 562.
Dryden 1936*b* *Aerodynamic theory* (Durand), 6, div. T.
Elias 1929 *Z. angew. Math. Mech.* no. 6 (December).
Elias 1930 *Z. angew. Math. Mech.* no. 1 (February).
Fishenden & Saunders 1932 *The calculation of heat transmission*, pp. 140–143. H.M.S.O.
Gilliland 1934 *J. Industr. Engng Chem.* **26**, 681 (January).
Himus 1929 Conference on vapour absorption and adsorption. Inst. Chem. Eng.
Hinchley & Himus 1924 *Trans. Instn Chem. Engrs*, **2**, 57–64.
Hine 1924 *Phys. Rev.* **24**, 79.
Hofmann 1932 *J. Industr. Engng Chem.* **24**, 135.
Jeffreys 1918 *Phil. Mag.* **35**, 278.

Kármán 1921 *Z. angew. Math. Mech.* **1**, 233.
Kármán 1934 *J. Inst. Aero. Sci.* **1**, 1–20.
Lettau 1939 *Atmosphärische Turbulenz*, p. 90.
Millar 1937 *Canad. Meteor. Mem.* **1**, no. 2.
Nikuradse 1932 *Forschungsh. Ver. dtsch. Ing.* no. 356.
Powell 1942 *Trans. Instn Chem. Engrs*, **18**, 136.
Powell & Griffiths 1935 *Trans. Instn Chem. Engrs*, **13**, 175.
Prandtl 1925 *Z. angew. Math. Mech.* **5**, 136.
Prandtl 1927 *Göttinger Ergebnisse*, **3**, 1.
Schiller 1923 *Z. angew. Math. Mech.* **3**, 2–13.
Sheppard 1940 *J. Sci. Instrum.* **17**, 218.
Skljarenko & Baranajew 1936 *Acta Physicochimica, U.S.S.R.* **4**, 873–888.
Sutton, O. G. 1934 *Proc. Roy. Soc.* A, **146**, 701–722.
Sverdrup 1936 *Geofys. Publ.* **11**, no. 7.
Taylor 1922 *Proc. Lond. Math. Soc.* **20**, 196.
Wade 1942 Paper read before a meeting of the Institute of Chemical Engineers (January).
Watson 1922 *Bessel functions*. Camb. Univ. Press.

The ultra-violet spectra and electron configuration of HgF and related halide molecules

By H. G. Howell, *University College, Southampton*

(*Communicated by W. E. Curtis, F.R.S.—Received* 11 *January* 1943)

[Plate 1]

The high-frequency ultra-violet emission spectrum of HgF has been photographed and analysed into two systems which are shown to be due to a $^2\Pi_r$–$^2\Sigma$ transition between Hg atom-like levels.

The constants are as follows:

	ν_e	ω_e	$x_e\omega_e$
$^2\Pi_{\frac{1}{2}}$	42990·6	469·4	10·05
$^2\Pi_{\frac{3}{2}}$	39060	506	—
$^2\Sigma$	0	490·8	4·05

The doublet separation is related to the ^2P width of the Hg atom, and the same separation is found in the spectra of the other Hg halides. It is shown that corresponding systems of the Zn and Cd halides should exist with doublet separations similarly related to the atomic ^2P level. Most of these have been identified amongst existing data, and as a result analyses of the bands of CdF, CdCl and ZnBr have been given although there is some evidence that the data attributed to ZnBr really belong to ZnI. In addition a system of CdI different in type from those above is analysed. A probable $^2\Sigma$–$^2\Sigma$ system overlapping the main $^2\Pi$–$^2\Sigma$ system is found for most of the molecules. All these systems together with those arising from the resonance line ^1P–^1S of the atom are discussed in terms of electron configurations of the states involved and an energy level diagram is presented.

Introduction

The original aim of this work was to analyse the spectrum of HgF in order to compare its vibrational constants with those of the members of such a series of heavy fluoride molecules as HgF, TlF, PbF and BiF, in which each one has one

more electron than the preceding. By following the changes in these constants from molecule to molecule it was hoped to gain information of their electron configurations. Data on the last three members of this series have already been published (Howell 1936; Rochester 1936), but nothing is known about the molecule HgF although papers have appeared on the other halides of mercury and also on the halides of zinc and cadmium, all of which should exhibit similarity of electronic constitution and spectra. All show extensive visible and ultra-violet band systems, the latter being rather difficult to analyse. This first led Wieland (1929) to believe that certain of the ultra-violet systems were due to a triatomic molecule. However, Cornell (1938) in his work on the ultra-violet spectra of the chlorides of Hg, Zn and Cd ascribes all these spectra to the diatomic molecule, and in his later work on HgCl Wieland (1941) accepts this correction. Cornell made the important observation that all the ultra-violet spectra studied by him occur near resonance lines of the metal atom Thus HgCl has two systems 2400–2650 and 2700–2900 with the 0,0 band of the former at 2516 (resonance line Hg 2537), a CdCl system occurs at 2185–2240 (resonance line Cd 2288) whilst ZnCl has two systems at 2076 and 2905–2980 (resonance lines Zn 2139 and 3076) This fact enabled him to correlate the various molecular states with probable products of dissociation.

The first half of this paper deals with the analysis of the HgF spectrum This has so many interesting features which yield information about the nature of the molecular states involved that it has been found possible to extend the scope of the investigation so as to include a critical review of all the halides of this group The later part of the paper deals with this attempt to systematise these spectra. As a result it is possible to check and correct, if necessary, previous work on these molecules and in some cases an interpretation of certain band systems, hitherto unanalysed, is put forward with some confidence.

Experimental

The emission spectrum of HgF has been developed in a high-frequency discharge using the apparatus and method described fully in former papers (Howell 1936). Details of spectrographs and plates used are as follows.

spectrograph	dispersion	plates	exposure time
E 1	2·5–2·8 A/mm.	Ilford Monarch and Double Xpress	30 min.
intermediate E 2	10 A/mm.	do.	2–5 min.

Fe arc lines were used as wave-length standards as well as the many Hg and Si lines found in the discharge, the weaker Hg lines in particular serving as excellent internal standards. The wave-length values used are those given in the *M.I.T. Tables* (1939). The accuracy of the wave-length measurements varies somewhat across the spectrum because of the nature of the bands to be described and also because of the difference in dispersion between the two spectrographs used. For the

TABLE 1. VIBRATIONAL ANALYSIS OF 2330 BAND SYSTEM OF HgF

v' \ v''		0		1		2		3		4		5		6		7		8		9	
0	Q	42987·1	481·5	42505·6																	
	P	42974·5																			$\Delta G'$
	QQ	449·6		449·9																	449·8
1	Q_1	43436 7	481·2	42955·5	474·3	42481 2	468	42013													
	P	*		42941 5	471 1	42470 4															
	Q_2	43471																			
	QQ			427·1		429·2		428													428 2
2				43382·6	472·2	42910 4	469 8	42440 6	460	41981											
						42891	467	42424 2	450	41974											
	QQ					408·7		409 8		410											409·5
3	Q					43319 1	468 7	42850 4	459 4	42391 0	447	41944									
										42370 3	438	41932									
	QQ					404		398		398 2		393									398·1
4						43723	475	43248	459	42789 2	457 0	42337 2	439	41898							
												42312 8	·	41887							
								405		396		394·6		392							394
5								43653	468	43185	453	42731 8	442	42290	432	41858					
												393		391		387					392
6												43125	444	42681	436	42245					
																395					
7																42640	429	42211			
																		398			
8																		42609	428	42181	
																				393	
9	$\Delta G''$		481 4		473·3		468 8		459·7		452·3		442		434					42574	
	(Q heads)																				

* Covered by Hg 2302

(*Facing p. 97*)

sharp heads measured on the E 1 the wave-lengths are considered accurate to 0·05 A or better. Because of the unfavourable wave-length—wave-number ratio in the region studied (2300–2600)—this amounts to 1 cm $^{-1}$. It is difficult to assess the maximum error of the diffuse or headless bands.

Description

A reproduction of the spectrum is given on plate 1. There is an apparent group of lines from 2300 to 2390 which on closer inspection turn out to be bands, mostly violet-degraded, with very short branches. Some are double-headed, some are like very sharp lines, whilst others are like diffuse doublets. The bands fall into sequences which degrade to the red, an unexpected direction for violet-degraded bands. Another interesting feature is that the sequence differences in a given sequence increase to a maximum and then decrease so that in one case a head is nearly formed. There is no trace of rotational structure in any of the bands

Around the resonance line 2537 there are apparently five strong violet-degraded bands with resolved structure which for such a heavy molecule as HgF cannot be rotational in nature. It is considered therefore to be sequence structure of bands with very short branches About eight bands can be detected in the stronger sequences. These are not sharp but appear as rather diffuse lines. In addition to these HgF systems there are Hg_2 bands present, in particular, the strong one at 2482 which has a similar close-sequence structure

Vibrational analysis

(*a*) 2330 *system*

The strongest bands in this system are around 2330, and these are taken to be Q heads of the $\Delta v = 0$ sequence with the fainter satellite heads on the long-wave side the corresponding P heads With this as basis the quantum scheme shown in table 1 has been drawn up The wave numbers of the bands of the $\Delta v = 0$ and -1 sequences are certainly the most accurate, whilst those belonging to $\Delta v = 1$ are the least reliable because of the diffuse and headless nature of the bands, with the exception of the first member which has probably a P and two Q heads which are degraded in opposite directions The P head is actually covered by the Hg line 2302, but inspection of the other members of this sequence makes it almost certain that it is present and comparatively sharp The weakest sequence $\Delta v = 2$ was studied with the Hilger E 2 spectrograph, and the few heads given here are due to red-degraded Q branches.

Inspection of the $\Delta G'$ values indicates a sudden reduction in $x'\omega'$ at about $v' = 4$ or 5. The apparent negative value of $x'\omega'$ after this is not considered real, as the small number of bands concerned does not afford opportunities for checking, and the measurements upon which the later $\Delta G'$ values are based are among the least accurate. It is probable that $x'\omega'$ becomes a steady constant value much smaller

than for the first vibrational levels. This discontinuity suggests the presence of a perturbing state and also rules out the possibility of representing all the bands in one quantum equation even with the addition of a cubic term. The following formula has been derived for the main unperturbed bands:

$$\nu_{(Q\ \text{heads})} = 42999{\cdot}6 + 469{\cdot}4u' - 10{\cdot}05u'^2 - 490{\cdot}8u'' + 4{\cdot}05u''^2,$$

where $u = v + \frac{1}{2}$.

The fit of the observed values to this formula can be examined in table 9. The nearness of the ω values accounts for the close sequence structure of the system and also for the headless nature of some of the bands A graphical analysis shows that the remaining bands of higher v' levels have an apparent ω'_e value of 410 cm $^{-1}$, and an $x'_e\omega'_e$ value of 1·5 cm.$^{-1}$. This discontinuity in the $x'\omega'$ is responsible for the uneven nature of the sequence difference. Thus a band in the $\Delta v = 0$ sequence is given by

$$\nu = \nu_{0,0} - (\omega'' - \omega')\,v'' - (\omega'x' - \omega''x'')\,v''^2.$$

Here, at first, $\omega'x' > \omega''x''$ and, as $\omega'' > \omega'$, the series in v'' is divergent and the bands in the sequence get farther apart. After $v' = 5$, however, $\omega'x' < \omega''x''$ and the series converges as do the band heads of the later members. As $\omega'' > \omega'$ the sequences degrade to the red, but as has already been pointed out the individual bands are not shaded in this direction. To account for this the full expression for the Q branch lines must be considered

$$\nu = \nu_0 + (B' - B'').J.(J+1) - (D' - D'').J^2.(J+1)^2.$$

Normally the D values are unimportant in determining the direction of shading of the branch as they are so much smaller than the B values However, in cases where $B' \doteq B''$ the heads are formed at high J values, and the contribution from the D terms may be considerable, even being so great as to reverse the sign of the expression as in the present case and so reverse the direction of degradation

(*b*) *The* 2560 *system*

The apparently violet-degraded bands in this system are interpreted as sequences which can be resolved into individual bands (apart from one sequence) which appear as coarse lines. The spacing of these bands is irregular at the head of each sequence, but after four or five they become more regularly spaced, closing up until they merge into each other. Because of this initial confusion of crowded bands at the head of each sequence it was at first thought that all that could be done with this system was to estimate the ω_e values from measurements on these heads. This showed that ω''_e was of the same order as that of the previous system. Assuming the similarity to extend to the type of electron transition, the initial confusion of bands can be explained by the presence of weak P branches which rapidly decrease in intensity down the sequence leaving only the Q branches as happens in the 2330 system. This is supported by the fact that the first member of each of the stronger sequences is weak. Thus measurement of the regular differences enabled an estimate of the sequence

difference to be made, this being of the same order as the difference in ω_e values Now the congestion of the bands in each sequence indicates that the ω_e values are very close together whilst the direction of sequence degradation shows that ω'_e is slightly greater than ω''_e. The average sequence difference is about 15 cm $^{-1}$, and combining this with the assumed value of 491 for ω''_e a value of 506 is obtained for ω'_e. The resulting quantum scheme for this system is given in table 2. All that can be inferred about $x'_e\omega'_e$ is that it is obviously small Using these ω_e values, an estimated value of 39060 cm.$^{-1}$ is found for ν_e

Electron transition

The similarity in vibrational constants of the two systems suggests that each is a component of a doublet system. The P and Q branches indicate that a $^2\Pi$ state is involved, and the other state will be shown later to be the $^2\Sigma$ ground state. This gives 3940 cm.$^{-1}$ as the doublet separation of the upper $^2\Pi$ state A comparison of this separation with the overall width of the ^{3}P state of the Hg atom brings out an important relation between them In Hg $6s6p$, ^{3}P the coupling constant a for the $6p$ electron is $\frac{2}{3} \times 6398 = 4265$ cm $^{-1}$ The corresponding molecular constant A for a π electron giving rise to a $^2\Pi$ state is given by $A = \Delta\nu = 3940$ cm.$^{-1}$ for HgF, i e it is of the same order as for the $6p$ electron. This throws some light on the electron configuration of the excited state and also shows that the excited electron is an almost unchanged Hg electron. Furthermore, the association of the system with the Hg resonance line 2537 shows that the transition is similar to $6s6p$, ^{3}P–$6s^2$, ^{1}S, i.e. it is atomic in nature. This view is further supported by the absence of change in the vibrational constants during the transition which is, therefore, essentially between non-bonding or atomic orbitals It follows then that little or no change in the doublet splitting should occur if a chlorine or other halogen atom replaces the fluorine atom in HgF The spectrum would be expected to be merely shifted to the red and to be characterized by lower vibrational coefficients Hence it is probable that unexplained features of other Hg halide spectra may be cleared up by examining them for two systems with a doublet separation of about 3900 cm $^{-1}$.

(*a*) HgCl

Both ultra-violet systems of this molecule have P and Q heads and therefore may form the $^2\Pi$–$^2\Sigma$ transition. Of the two systems that between 2400 and 2650 is the simpler, and the analysis given by Cornell is undoubtedly correct. He also gave an analysis of the 2700–2900 system in which he stated that because of slight differences in the values of the vibrational constants the lower state was not common to both systems Wieland has since expressed doubt as to the correctness of this latter analysis and claims that the 0, 0 band must be at 2790 or 2812 instead of at 2741. He bases this on a study of the spectrum of $HgCl_{37}$ but gives no details. Cornell observed that his bands were accompanied by weaker ones displaced by amounts compatible with them being ascribed to the isotopic molecule $HgCl_{37}$. Such an ex-

TABLE 2. VIBRATIONAL ANALYSIS OF 2560 SYSTEM OF HgF

v' \ v''	0		1		2		3		4		5		6		7		8		9
0	39053	479	38574 490	479	38095 490														
1			39064 497	479	38585 493	475	38110 492												
2			39561* 494	483	39078 495	476	38602 491	477	38125 496										
3			40055†	482	39573	480	39093 494	472	38621 486	479	38142 496								
4							39587	480	39107 491	469	38638 478	483	38155 493						
5									39598	482	39116 488	468	38648 485	480	38168 493				
6											39604	471	39133 486	472	38661 482	477	38184 492		
7													39619	476	39143 484	477	38676 485	479	38197 477
8															39627	466	39161	477	38684 491
9																			39175

* This sequence may need to be moved up one member. Although present numbering gives better ΔG values, this is not considered to be certain, as the error in measurement is of same order as the sequence difference.

† Rest of sequence unresolved.

planation, however, demands the location of the system origin to be at the red end which certainly is unusual, and so he discarded this view in favour of that which ascribes the weaker bands to P branches. This is consistent with the system being a $^2\Pi$–$^2\Sigma$. Now the origin of the shorter wave-length system is at 39704 cm $^{-1}$, and assuming the same doublet separation as for HgF the origin of the other component should be at 35760 cm.$^{-1}$, corresponding to a wave-length of 2796 A which is so close to the two chosen by Wieland that it is tempting to consider the question settled. If Cornell's origin is correct the doublet separation would be 3200 cm $^{-1}$, a value which, if small, is nevertheless of the expected order. Examination of his reproduction of the spectrum reveals the following facts:

(1) The bands analysed by Cornell are arranged in very close sequences of headless bands almost identical in appearance with the longer wave system of HgF

(2) There are other bands present, the strongest one being about 2790 and others at 2749 and 2811. The first is actually the strongest band in the whole photograph.

These bands are mentioned by Cornell, but he omitted to point out an important difference between these and those analysed by him, viz that they are without sequence structure, the individual bands being quite unresolved. This suggests that there are two systems present, the unresolved bands belonging to one in which the branches are longer than in the other. Intensity considerations indicate the unresolved sequence at 2790 to be the $\Delta v = 0$, whilst the spacing of the other sequences show that the vibrational constants are almost identical with those of the Cornell system Wieland has not yet published details of his work on $HgCl_{37}$, but if his view that the double heads of this system are isotope heads is correct, then there is no need to assume that a $^2\Pi$ state is involved. An examination of the relative intensities of the so-called P and Q heads would settle this, for with HgF the P heads in a given sequence rapidly decrease in intensity relative to the Q heads. These details cannot unfortunately be seen in his plate. At present it will therefore be assumed that the band at 2790 is the 0, 0 band of the $^2\Pi$ $^2\Sigma$ system, this giving about 3900 cm.$^{-1}$ as the doublet separation.

(*b*) HgBr

This molecule has an extensive ultra-violet spectrum, part of which has been analysed by Wieland who identified a system origin at 38574 cm.$^{-1}$. The long-wave end of the spectrum was first considered by him to be due to a polyatomic molecule because of its complexity, but it is now realized that this is unlikely. The strongest groups of heads listed are those at 34580 and 34668 cm.$^{-1}$, and it is reasonable to assume that one of these is the first of the $\Delta v = 0$ sequence of the other $^2\Pi$–$^2\Sigma$ component. This gives a $^2\Pi$ separation of either 3996 or 3906 cm.$^{-1}$, both being in remarkable agreement with the values for HgF and HgCl It may well be that there are two systems here as seems to be the case with HgCl.

(*c*) HgI

There are many bands of this molecule in the expected region of the $^2\Pi$–$^2\Sigma$ transition, and according to Wieland (1932) they can be analysed into two systems having a common lower state with $\omega'' = 125$ cm.$^{-1}$. The ν_e values are 32729 and about

36100 cm.$^{-1}$ which gives a difference of 3400 cm.$^{-1}$, a reasonable indication that these bands do belong to this transition, but the $^2\Pi$ separation does not compare quite so favourably with those of the other halides so that it may be that one of the origins is out by as much as 500 cm.$^{-1}$.

Other systems have been reported by Prilheshejewa (1932) from 2000 to 2300 and Sastry (1941) around 2600. The latter presents an analysis of some of these bands into a $^2\Pi$–$^2\Sigma$ system with a lower state $\omega'' = 55{\cdot}5$ cm $^{-1}$ and a doublet separation of 766 cm.$^{-1}$ Another $^2\Pi$–$^2\Sigma$ system is considered to occur around 2540 with a lower state ω'' of 92·7 cm.$^{-1}$ and a doublet separation of 126 cm $^{-1}$. Although few details are given by this writer it must be pointed out that none of the vibrational frequencies of his systems refer to the ground state nor to the upper state of Wieland's bands. Furthermore, it may be significant that the 126 cm.$^{-1}$ separation is so close to the ground state ω.

Application to the halides of Zinc and Cadmium

The proximity of the chloride spectra of Zn and Cd to the metal resonance lines noticed by Cornell suggests a probable similarity in electron transition to that of the Hg halides. Thus the spectrum of the Zn halides resulting from the excitation of a Zn 4*s* electron should be due to a $^2\Pi$–$^2\Sigma$ transition in which the $^2\Pi$ separation is $\frac{3}{2}(^3\mathrm{P}, \mathrm{Zn}) = 386$ cm.$^{-1}$. The corresponding characteristic doublet separation of the Cd halide spectra will be $\frac{3}{2}(^3\mathrm{P}, \mathrm{Cd}) = 1140$ cm.$^{-1}$. If, then, a doublet system is found with this separation for one of the halides, the same doubling will almost certainly occur for all the others. With this in mind, together with the fact that corresponding systems for fluoride, chloride, etc. occur progressively towards longer wave-lengths, the way to interpret existing data seems clear.

(*a*) *Zinc*

Rochester (1939) has found that the 2670–2700 system of ZnF is a doublet with a separation of 370 cm.$^{-1}$ in excellent agreement with the predicted value. Better still is the 383·5 cm.$^{-1}$ separation found by Cornell as the doublet splitting of the 2905–2980 system of ZnCl. There is thus no doubt that these systems are the counterparts of those of Hg, and furthermore they can be confidently looked for in the spectra of ZnBr and ZnI.

With ZnBr, Walter & Barratt (1929) have reported bands in absorption from 3027 to 3113. Their strongest bands are at 3071 and 3110, and assuming these to be the 0, 0 bands of the $^2\Pi$–$^2\Sigma$ system a splitting of 408 cm.$^{-1}$ results. With this as starting-point, it has been found possible to fit all the bands listed by these workers into the quantum scheme shown in table 3. Although they give no information about the nature of the bands, it is probable that they are of the short-branched and often headless type, similar to the HgF bands of the 2560 system. In any case the wave-lengths are only given to the nearest Angstrom, and so the irregular ΔG values, also found with HgF, should not be considered as throwing doubt on the correctness of this analysis. Strong support is found in the intensity distribution which is

TABLE 3. VIBRATIONAL ANALYSIS OF ZnBr BANDS*

Q heads

v'' \ v'	0		1		2		3		4		5		6
0	32553(4)	246	32799(0) 214										
1			32585(3)	257	32842(0) 214								
2					32628(2)	268	32896(0) 226						
3							32670(1)	269	32939(00) 226				
4									32713(0)	270	32983(00) 206		
5											32777(00)	249	33026(00)

System II consists only of the $\Delta v = 0$ sequence

32145(4)–32228(2)–32322(1)–32405(1)
32114(1P)

* These bands may really belong to ZnI. Intensities given by Walter & Barratt.

TABLE 4. PROBABLE CLASSIFICATION OF CdF *Q* HEADS

	0, 0		0, 1		0, 0
	34930(5)	552	34378(2)		33757(5)
	533		34425(1*R*)		489
1, 0	35463(3)			0, 1	33268(5)
	System I				System II

characterized by strong $\Delta v = 0$ sequences flanked by much weaker sequences. Thus, although system II is represented by only the one sequence it is considered certain that longer exposures would bring up the others. Approximate values of the vibration frequencies are $\omega' = 250$ and $\omega'' = 220$ cm $^{-1}$. An unexpected complication arises when the next molecule in the series is examined. ZnI has only one definitely established system, and Wieland's analysis gives 3318 as the 0, 0 band with vibrational frequencies almost identical with those derived above for ZnBr. The order of magnitude of these values is that to be expected for ZnI, and so it may be suspected that the bands attributed by Walter & Barratt to ZnBr may really be due to the iodide The bands are certainly in the position expected for the bromide, but the only way to resolve the difficulty is to assume that the stronger sequence in table 3 is not $\Delta v = 0$ but $\Delta v = -1$ Whilst this will give ω values of the right order, such an intensity displacement would be difficult to understand There is obvious need here of a new investigation, but in the meantime the bands will temporarily be allocated to the bromide In his study of ZnI Wieland also mentioned the occurrence of a weak group of heads at 3236. It is unlikely that this group forms the nucleus of a $^1\Pi$–$^1\Sigma$ system, as this gives a doublet splitting of 760 cm.$^{-1}$. Assuming a normal value of 400 cm.$^{-1}$ for this, a predicted system origin should be about 3276, if it is accepted that 3318 is the origin of the longer wave-length component.

(*b*) *Cadmium*

Bands in the visible, reported by Asundi, Samuel & Zaki-uddin (1935) to be due to CdF, have been identified by Pearse & Gaydon with CaF. The latter workers give in their *Identification of molecular spectra* (1941) the wave-lengths of six bands obtained in an arc between Cd electrodes fed with cadmium fluoride, but point out that their assignment to CdF is uncertain It is now possible to show that they are certainly due to this molecule. Examination of the value for the shorter-wave system of the $^1\Pi$–$^1\Sigma$ transition shows that for ZnF and HgF they occur 4840 and 3590 cm.$^{-1}$ respectively, deeper than the ^{3}P–^{1}S resonance line. Since Cd occurs between Zn and Cd in the atomic scheme, a mean value of 4200 cm $^{-1}$ can be assumed for the corresponding CdF quantity. Thus, adding this to 30660 cm $^{-1}$ the wave number of the Cd ^{3}P–^{1}S line gives 34860 cm.$^{-1}$ or 2868 A as the location of a system origin. With a doublet separation of 1140 cm.$^{-1}$ the second system should occur at 2965 A. Now two of the three strongest bands listed by Pearse & Gaydon are 2961·5 and 2862 A, and the nearness of these to the predicted values cannot be accidental but must surely mean that they are indeed the 0, 0 bands of the doublet system. This leads to an actual doublet separation of 1179 cm.$^{-1}$. It is now possible to arrange all the bands in a quantum scheme which is shown in table 4. From this it appears that a probable value of ω'' is about 550 cm.$^{-1}$, a value of the expected order, as it should be intermediate between 620 cm.$^{-1}$ for ZnF and 475 cm $^{-1}$ for HgF. The scheme is further supported by the intensity distribution

No CdCl system in the anticipated region was observed by Cornell, but an analysis of a brief system between 2185 and 2240 yielded an ω value of 330·5 cm.$^{-1}$, and this is probably the ground-state frequency. Walter & Barratt found absorption bands

between 3181 and 3018, and their strongest two listed are at 3181 and 3072 Assuming as usual that these head separate $\Delta v = 0$ sequences a doublet separation of 1115 cm.$^{-1}$ results, a value which compares so favourably with the two previous ones that some confidence must be placed in the following quantum analysis (table 5). Approximate ω values are $\omega'' = 330$ cm.$^{-1}$, $\omega' = 400$ cm.$^{-1}$, the former being in excellent agreement with that obtained by Cornell for the other system. As with other cases one system is very poorly represented, but the intensity distribution affords further support for the analysis suggested. All the bands have been classified assuming that some heads can be interpreted as R heads, these being expected in a $^2\Pi$–$^2\Sigma$ transition.

Wieland has analysed one system of CdBr with the 0, 0 band at 3177 and lists three bands which belong to another system. The strongest of these lies at 3095, i.e. 840 cm.$^{-1}$ away. This is too small to be the doublet separation, and besides, comparison with CdCl suggests that the missing system should lie on the long-wave side of that at 3177, i e. it can be expected to occur about 3296

With CdI the same worker has reported a system at 3384 (0, 0 band) and three unidentified bands at 3586, 3563 and 3541, the last being the strongest. If it is a 0, 0 band the doublet spacing is 1300 cm.$^{-1}$, which is of the right order. Furthermore, the intervals between these bands are those of the ground state (178·5 cm.$^{-1}$). However, the interval of 1300 cm.$^{-1}$ is not considered sufficiently close to the anticipated value of 1140 cm.$^{-1}$ to justify the definite allocation of these bands to the $^2\Pi$–$^2\Sigma$ system.

To summarize these results, the actual or probable 0, 0 bands of all the $^2\Pi$–$^2\Sigma$ transitions are collected in table 6

Other ultra-violet systems

In the cases of CdCl and ZnCl brief systems were found by Cornell to occur on the short-wave side of the metal resonance line ^{1}P–^{1}S, to which they are obviously related as the $^2\Pi$–$^2\Sigma$ systems are to the ^{3}P–^{1}S lines These are considered by him to be $^2\Sigma$–$^2\Sigma$ systems No such system was found in the case of HgCl in the neighbourhood of Hg 1850, but it is probable that this lies too far in the ultra-violet for normal detection These $^2\Sigma$–$^2\Sigma$ systems should also be present for the other halides under discussion, the bromides and iodides being displaced more and more to the red. This means that these latter molecules will have their resonance system in a more accessible region of the spectrum. In particular, CdI should have its system well placed for photographing. Wieland lists a number of bands from 2360 to 2550 for CdI in his table 15, and, assuming that some of these heads are double, they can all be arranged in the quantum scheme given in table 7. The doubling is tentatively attributed to the presence of Q and R branches, the relative intensities being in keeping with such an interpretation. The analysis is supported by the intensity distribution, this being an open parabola characteristic of all iodides and also by the magnitude of the ω values, viz $\omega' = 106$ cm.$^{-1}$, $\omega'' = 180$ cm.$^{-1}$, the former being in agreement with that found by Wieland for the $^2\Pi$–$^2\Sigma$ system.

It is probable that an analogous system will be found for HgI among the bands already reported by Sastry and by Prilhesheiewa

TABLE 5. VIBRATIONAL ANALYSIS OF CdCl BANDS

System I (Q heads)

v' \ v''	0		1		2		3		4		5
0	32543 (5) 32521 ($2P$) 386	336	32207* (0) 32249 ($0R$) 399	308	31899* (0) 381						
1	32929 (0)	323	32606 (4) 388	326	32280* (0) 390	310	31970* (0) 383				
2			32994 (0)	324	32670 (3) 400	317	32353* (0) 381	311	32042* (00) 385		
3					33070 (0)	336	32734 (2) 391	307	32427* (0) 383	334	32093* (00) 397
4							33125 (0)	315	32810 (1)	320	32490* (0) 374
5											32864 (1)

System II (Q heads)

Confined to $\Delta v = 0$ sequence:

31428 (5) 31517 (5) 31606 (3) 31717 (1) 31787* (0)
31497 ($2P$)

In brackets, intensities according to Walter & Barratt.

* Difficult to measure.

TABLE 6. THE 0, 0 BANDS OF THE $^2\Pi$-$^2\Sigma$ SYSTEM

	Zn	Cd	Hg
F	2677–2704	2862–2962	2326–2560
Cl	2942–2976	3072–3181	2516–2790
Br	3071–3110*	3177–3296†	2592–2891 or 2883
I	‡	3384–3520§	2769–3055‖

* This doublet may refer to ZnI.
† 3296 is estimated.
‡ Data discussed in text are not given here because of *.
§ 3520 is estimated.
‖ 2769 is estimated.

TABLE 7. VIBRATIONAL ANALYSIS OF CdI BANDS

v'' \ v'	0		1		2		3
0							42288·0(0)
							176·0
1					42009·5(1)	102·5	42112·0(1)
					175·0		
2			41732·0(1)	102·5	41834·5(3)	102·5	41937 0(2)
			171·5		157·0		173 5
3			41560·5(2)	100 5	41661·0(3)	102·5	41763·5(2)
			175·5		173·5		
4	41281·5(1)	103·5	41385·0(3)	103·0	41488·0(2)		
	169 0		174 0		173·0		
5	41112·5(1)	98·5	41211·0(3)	104 0	41315 0(1)		
			41217·0(3*R*)				
	173·0		165·0				
6	40939 5(3)	105·5	41046 0(2)				
	40943·5(1*R*)						
	171·5		170 5				
7	40768·0(2)	107 5	40875 5(1)				
	40775 5(1*R*)						
	168 0						
8	40600·0(2)				40812·5(0)		
	40607·5(1*R*)						

Intensities according to Wieland

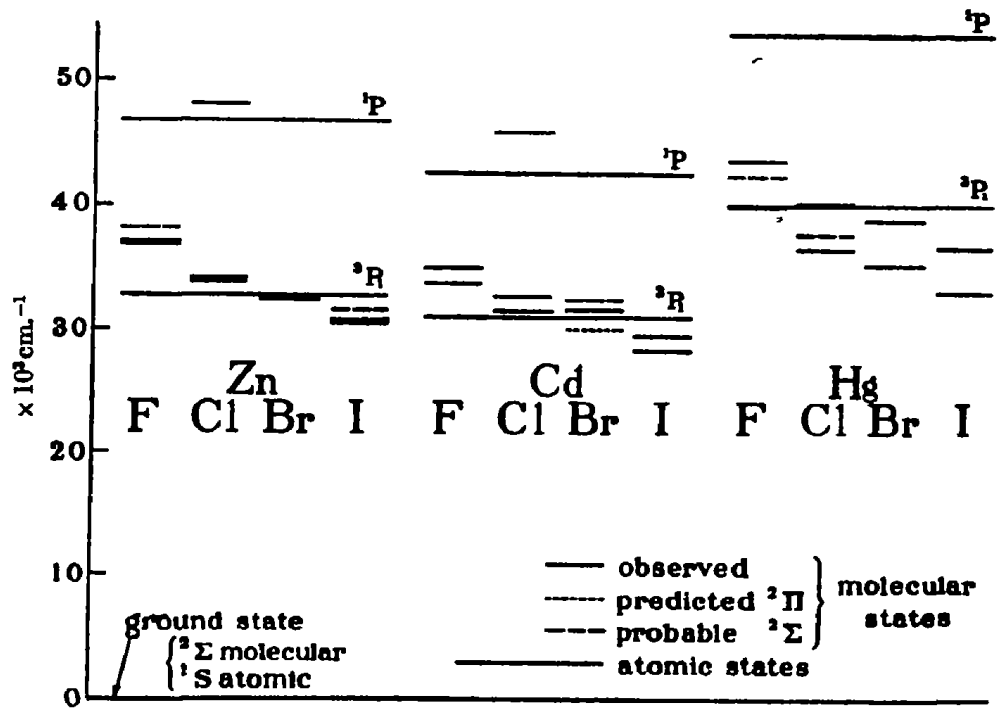


FIGURE 1

Besides this further ultra-violet system there is evidence of a system probably $^2\Sigma$–$^2\Sigma$ occurring near to the $^2\Pi$–$^2\Sigma$ system. As already mentioned there seems to be one mixed up with the 2700–2900 system of HgCl, and the perturbations discussed for HgF may well be due to the presence of a $^2\Sigma$ state. Also Rochester definitely found such a system in the case of ZnF, there being evidence of interaction between the $^2\Sigma$ and $^2\Pi$ upper states. Other bands not fitting into the main $^2\Pi$–$^2\Sigma$ system of ZnI, CdBr and CdI also possibly belong to such a $^2\Sigma$–$^2\Sigma$ system. In any case there is definite evidence of the existence of a $^2\Sigma$ state of about the same energy as the $^2\Pi$. All the states, known and predicted for all these halides, are collected in figure 1, with the exception of the $^2\Sigma$ state which is the upper level of the visible bands. This is well known for HgCl but has not yet been studied for any other molecule. The corresponding atomic levels are shown to demonstrate how closely the molecular levels follow them.

The position of the $^2\Sigma$ in HgF is estimated from the fact that perturbation of the $^2\Pi$ level sets in at about $v' = 5$

Electron configuration

The next task is to attempt to interpret the spectra in terms of electron configurations of the states concerned Information about the configuration of HgF can be obtained from earlier members of the same group, such as BeF, and also from other Hg molecules It is now accepted that the BeF ground-state configuration can be written $KK\sigma^2\sigma^2\pi^4\sigma$, $^2\Sigma$ and some idea of the nature of the closed shells in this structure can perhaps be best understood by referring to the molecule with one less electron BeO whose $^1\Sigma$ ground state has the same configuration minus the last σ electron. The addition of this electron to BeO to form BeF results in a reduction in ω_e from 1487 to 1266 cm.$^{-1}$, indicating that the added electron has a definite anti-bonding tendency The BeO state $\sigma^2\pi^3\sigma$, $^1\Pi$ shows a bigger reduction to 1128 cm.$^{-1}$, which can be explained by assuming that the σ electron is occupying the anti-bonding orbital found in BeF and that the π^4 group is very strongly bonding. Indeed, it is probable that most of the molecular binding is due to this group. Thus the state $\sigma\pi^4\sigma$, $^1\Sigma$ gives a slight reduction in ω_e to 1371 cm.$^{-1}$, showing that the inner σ^2 shell has little bonding power. The state $\sigma\pi^3\sigma^2$, $^1\Pi$, which results from the removal of a strongly bonding π from $\sigma\pi^4\sigma$, $^1\Sigma$ to the slightly anti-bonding σ orbital should result in a big reduction in ω_e. Actually the ω_e reduction is from 1371 to 1006 cm $^{-1}$, i.e is the same numerically (360 cm.$^{-1}$) as that in the transition $\sigma^2\pi^4$–$\sigma^2\pi^3\sigma$, which is not surprising as it represents essentially the same transition. Applying these ideas to BeF, probable facts about the first lowlying states can be inferred. These are:

ground state	$\sigma^2\pi^4\sigma$, $^2\Sigma$
(1)	$\sigma^2\pi^3\sigma^2$, $^2\Pi_i$
(2)	$\sigma\pi^4\sigma^2$, $^2\Sigma$

(1), which represents a transition from a strongly bonding to an anti-bonding orbital, should have a much lower ω than the ground state, whereas that of (2) should be only slightly less. An excited $^2\Pi$ state is known, but its nature is rather uncertain, and although ω_e' is less than ω_e'' the difference is nothing like that to be expected. In his rotational analysis of the $^2\Pi$–$^2\Sigma$ system Jenkins (1930) showed that the coupling constant A in the $^2\Pi$ state is either $= 22$ (normal) or 16·46 cm.$^{-1}$ (inverted). Because of insufficient resolution of lines having low J values it was not possible to decide between these two alternatives on the basis of the missing lines, but from Λ-type doubling measurements he tentatively concluded the $^2\Pi$ state to be normal, in which case the $^2\Pi$ must derive from such a configuration as $\pi^4\pi$. A more detailed consideration of Jenkins's data, however, led Mulliken (1931) to the opposite view, although he admitted that the situation is not satisfactory. For example, when compared with the isoelectronic molecules BO, CO and CN the smallness of the constant ($-16\,5$) is puzzling when compared with the expected value of 271, although it is much larger than the A derived from the ^{3}P state of Be, viz. $+2{\cdot}0$. As a possible way out of the difficulty he considered that the actual state may be a hybrid of $^2\Pi_r$ and $^2\Pi_i$ which would show both tendencies, and this could be brought about by the two states $\pi^3\sigma^2$, $^2\Pi_i$ and $\pi^4\pi$, $^2\Pi_r$ being so close in energy that by quantum interaction one or more $^2\Pi$ states intermediate in properties might result. Before inquiring how the structure of HgF compares with that of BeF or whether our knowledge of the spectrum of HgF can throw any light on to this problem, it is instructive to examine the hydrides of Hg, HgH^+ and HgH. Thus the transition $6s\sigma\,6p\sigma$, $^1\Sigma$–$6s\sigma^2$, $^1\Sigma$ in HgH^+ is essentially a $6s\,6p$–$6s^2$ Hg transition, as the electronic energy is roughly the same, viz. 44000 cm.$^{-1}$ Further, Mulliken (1932) has shown that the doublet spacing of 3700 cm.$^{-1}$ for the upper $^2\Pi$ state of HgH is due to a $6p\pi$ electron as with HgF, and so the system is due to a $6s\sigma^2\,6p\pi$, $^2\Pi$–$6s\sigma^2\,6p\sigma$, $^2\Sigma$ transition. HgF resembles both these molecules, for not only does the observed doublet separation accord with that expected for a $6p\pi$ electron, but the electronic energy shows that the transition is practically a $6p$–$6s$ Hg as with HgH^+ Thus it is evident that with HgF the transition is between a $6s\sigma$ and a $6p\pi$ electron and that therefore the actual configurations are $\sigma^2\pi^4p\pi$, $^2\Pi_r$–$\sigma^2\pi^4s\sigma$, $^2\Sigma$, i.e. a regular $^2\Pi$ upper state is to be expected. Comparing the ground state $\sigma^2\pi^4s\sigma$, $^2\Sigma$ arising from ^{1}S (Hg) and ^{2}P (F) with that of BeF, it is probable that the π^4 group again contributes most of the binding whilst the outer σ which in BeF is slightly anti-bonding is now non-bonding or atomic The first excited states derived from this ground state will be $\sigma^2\pi^3s\sigma^2$, $^2\Pi_i$ and $\sigma\pi^4s\sigma^2$, $^2\Sigma$. The first which represents a transition from a molecular bonding orbital to an atomic or non-bonding orbital should give a system in which $\omega_e' < \omega_e''$, but the known $^2\Pi_r$ state is not of this type. As a molecule gets heavier the outer shells differ less and less in energy so that the $^2\Pi_i$–$^2\Sigma$ system can be expected to occur in the infra-red. Returning to the problem of the BeF $^2\Pi$–$^2\Sigma$ system, it is now possible to suggest an explanation. It is definite that $\pi^4\pi$, $^2\Pi_r$ applies to HgF and that the evidence for BeF is unable to decide between the same configuration and $\pi^3\sigma^2$, $^2\Pi_i$. It is tempting to consider that when better data are available for BeF the $^2\Pi$ will be found to be regular. However, in spite of the fact that Be and Hg

have the same electron configuration there is an important difference between them. The spectra under discussion, of the type HgF, have been shown to bear a strong resemblance to the atomic ^3P–^1S transition. Now such an intercombination line is not found for Be, and in the other cases its intensity increases down the Periodic Column towards Hg. The transition probability for the atom will be retained to some extent in the molecule, especially if atomic orbitals are involved. Thus the HgF system should be the strongest whereas that of BeF should tend towards zero intensity in the absence of any disturbing factors. Such a factor is the presence of a perturbing state with which it can mix in the way suggested by Mulliken, so that a transition with the ground state may now be possible. Thus the uncertainty in the nature of the BeF state is probably fundamental.

The state $\sigma\pi^4 s\sigma^2$, $^2\Sigma$ may be the upper state of the visible bands. These have been studied by Wieland in the case of HgCl who finds $\omega_e' < \omega_e''$ as is to be expected Other excited states may be expected as follows: $\pi^4 p\sigma(^1P)$, $^2\Sigma$ and $\pi^4 p\pi(^1P)$, $^2\Pi_r$, these arising from ^1P (Hg), and so this symbol is included in the configuration to distinguish them from those which derive from ^3P (Hg). If Cornell's assumption that the further ultra-violet systems of ZnCl and CdCl are due to $^2\Sigma$–$^2\Sigma$ transitions is correct, then they are represented by $\pi^4 p\sigma(^1P)$, $^2\Sigma$–$\pi^4 s\sigma$, $^2\Sigma$. The location of the accompanying levels $\pi^4 p\pi(^1P)$, $^2\Pi$ and $\pi^4 p\sigma(^3P)$, $^2\Sigma$ will then depend upon the $p\pi$–$p\sigma$ separation, the $p\sigma$ being normally lower. In HgH this separation is 26800 cm.$^{-1}$, but there is no justification for assuming a similar value here. It is curious that although the $^2\Pi$–$^2\Sigma$ system occurs near to the ^3P–^1S resonance line, it is the $^2\Sigma$–$^2\Sigma$ system which is apparently near to the ^1P–^1S line. It is, however, not certain that the designation as $^2\Sigma$–$^2\Sigma$ is correct, dependent as it is upon the appearance of only a few bands Indeed, if the CdI bands analysed in this paper belong to the corresponding system, then the presence of the Q and R heads indicates that it is really a $^2\Pi$–$^2\Sigma$ transition. The similarity of the fragmentary $^2\Sigma$–$^2\Sigma$ system to the main $^2\Pi$–$^2\Sigma$ system, particularly with regard to the vibrational frequencies, may mean that the upper $^2\Sigma$ state can be identified as $\pi^4 p\sigma(^3P)$, $^2\Sigma$. In this case the $p\pi$–$p\sigma$ separation is very small and may even be negative.

Comparison of vibration frequencies

The ω values of all the fluoride molecules discussed at the beginning of this paper are shown in table 8.

Table 8 Ground state vibration frequencies

molecule	HgF	TlF	PbF	BiF

They are all of the same order, and a similar study of the ω values of the other halides shows that for the heavier diatomic molecules ω is largely determined by the molecular mass and is relatively insensitive to changes in electron configuration.

Classification of HgF Bands

All of the HgF bands measured are shown in the usual way in table 9. No intensity value is given for the bands of the 2560 system, as their headless nature renders such a task difficult and meaningless unless the integrated intensity of the whole band is measured.

Table 9. Classification of HgF bands

2330 system. Q heads

λA	I	ν cm.$^{-1}$	O–C	v′, v″	λA	I	ν cm.$^{-1}$	O–C	v′, v″
2286·4	0	43723		4, 2	2351·91	8	42505 6	0 9	0, 1
2290·1	0	43653		5, 3	2353 26	5	42481·2	1 8	1, 2
2299·7	4	43471		1, $0Q_1$	2353·86	1	42470·4		1, 2P
2301·49	4	43436·7	0 0	1, 0	2355 51	5	42440·6	−1·4	2, 3
2304·36	4	43382 6	−0 6	2, 1	2356 42	1	42424 2		2, 3P
2307·74	3	43319·1	1·4	3, 2	2358·27	4	42391·0	−1·6	3, 4
2311 5	2	43248		4, 3	2359 42	1	42370·3		3, 4P
2314 9	1	43185		5, 4	2361 27	3	42337·2		4, 5
2318 1	1	43125		6, 5	2362 63	0	42312 8		4, 5P
2325 56	10	42987 1	−0 3	0, 0	2363 9	2	42290		5, 6
2326·25	8	42974·5		0, 0P	2366·4	1	42245		6, 7
2327·27	8	42955 5	1·5	1, 1	2368 3	0	42211		7, 8
2327 94	6	42941 5		1, 1P	2370 0	0	42181		8, 9
2329 72	8	42910 4	1 8	2, 2	2379·5	3	42013	0	1, 3
2330·8	5	42891		2, 2P	2381·3	2	41981	−2	2, 4
2332 98	7	42850 4	−0 7	3, 3	2381 7	0	41974		2, 4P
2336 32	6	42789 2		4, 4	2383·4	2	41944	2	3, 5
2339 46	5	42731·8		5, 5	2384 1	0	41932		3, 5P
2342·2	4	42681		6, 6	2386·0	1	41898		4, 6
2344 5	3	42640		7, 7	2386 6	0	41887		4, 6P
2346 2	2	42609		8, 8	2388·3	0	41858		5, 7
2348·1	0	42574		9, 9					

2560 system. Q branch settings only

λA	ν cm.$^{-1}$	v′, v″	λA	ν cm.$^{-1}$	v′, v″
2495 8	40055	3, 1	2584·3	38684	8, 9
2522·8	39627	8, 7	2584·8	38676	7, 8
2523 3	39619	7, 6	2585 8	38661	6, 7
2524 2	39604	6, 5	2586·7	38648	5, 6
2524 6	39598	5, 4	2587·4	38638	4, 5
2525·3	39587	4, 3	2588·5	38621	3, 4
2526·2	39573	3, 2	2589 8	38602	2, 3
2527·0	39561	2, 1	2590·9	38585	1, 2
2551·9	39175	9, 9	2591 6	38574	0, 1
2552·8	39161	8, 8	2617·2	38197	7, 9
2554·0	39143	7, 7	2618 1	38184	6, 8
2554·6	39133	6, 6	2619·2	38168	5, 7
2555·8	39116	5, 5	2620·1	38155	4, 6
2556·3	39107	4, 4	2621·0	38142	3, 5
2557·3	39093	3, 3	2622·2	38125	2, 4
2558·2	39078	2, 2	2623·2	38110	1, 3
2559·1	39064	1, 1	2624·2	38095	0, 2
2559·8	39053	0, 0			

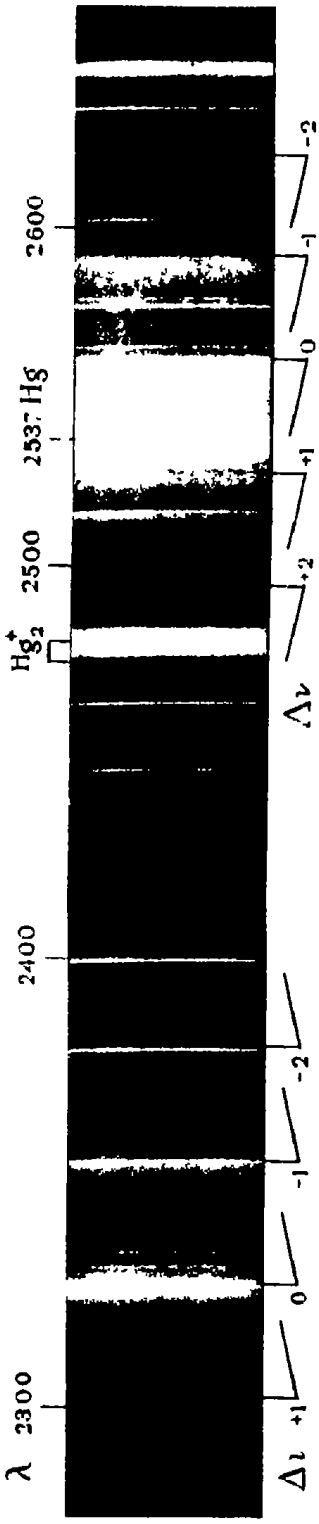
I am grateful to Professor W. E. Curtis, F.R.S., for the experimental facilities placed at my disposal at King's College, Newcastle.

Note on CdF, *added in proof.*

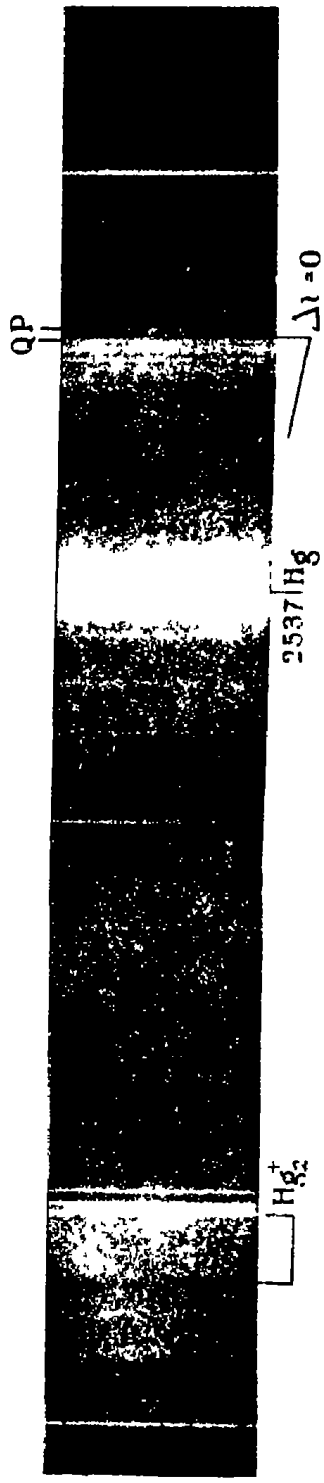
My attention has been drawn to a recent publication by Fowler (*Phys. Rev.* **62**, 141, 1942) which was not available when the present paper was prepared. By heating CdF_2 from 1350 to 1600° C he obtained absorption bands, attributed to CdF of which six form a progression and three belong to an isolated sequence. Each of the six bands is apparently really a sequence of heads, unresolved because of the near equality of the ω values which are somewhat larger than 535 cm.$^{-1}$. These bands all occur in the same wave-length region as those of Pearse and Gaydon (P & G) but none coincide. Fowler states that the ΔG values found by him do not occur in the bands of P & G, but it is shown in the present paper that the data of P & G provide a probable value of 550 cm.$^{-1}$ for ω_e''. The similarity of the ΔG values makes it probable that both absorption and emission bands are in fact due to CdF. The main bands obtained by Fowler are probably due to the $^2\Sigma$–$^2\Sigma$ transition found in so many of the analogous molecules to be situated in exactly the same region as the $^2\Pi$–$^2\Sigma$. That three of the six bands have satellite heads is no objection, as the corresponding $^2\Sigma$–$^2\Sigma$ system in CaF and SrF, interpreted by Harvey (*Proc. Roy. Soc.* A, **133**, 336, 1931), have similar satellite heads and a comparison of spectrograms shows that the relative intensity of main to satellite heads is about the same as with Fowler's bands. It would appear then that the $^2\Sigma$–$^2\Sigma$ transition has a greater probability in absorption than $^2\Pi$–$^2\Sigma$, but the use of higher temperatures should bring up the bands of P & G. A more complete study of the emission spectrum is obviously desirable.

References

Asundi, Samuel & Zaki-uddin 1935 *Proc. Phys Soc* **47**, 235.
Cornell 1938 *Phys. Rev.* **54**, 341.
Howell BiF 1936 *Proc. Roy. Soc.* A, **155**, 141.
Howell TlF 1937 *Proc. Roy. Soc.* A, **160**, 242.
Jenkins 1930 *Phys. Rev* **35**, 315.
M.I.T. Wavelength Tables 1939.
Mulliken 1931 *Phys. Rev.* **38**, 836.
Mulliken 1932 *Rev. Mod. Phys.* **4**, 1.
Pearse & Gaydon 1941 *Identification of molecular spectra*, p. 97. London: Chapman and Hall.
Prilhesheiewa 1932 *Phys. Z. Sowjet.* **1**, 189.
Rochester PbF 1936 *Proc. Roy. Soc.* A, **153**, 407
Rochester 1939 *Z. Phys.* **114**, 495.
Sastry 1941 *Curr. Sci.* **10**, 521.
Walter & Barratt 1929 *Proc. Roy. Soc.* A, **122**, 201.
Wieland 1929 *Helv. Phys. Acta*, **2**, 46.
Wieland 1932 *Z. Phys.* **76**, 801.
Wieland 1941 *Helv. Phys. Acta*, **14**, 420.



Emission spectrum of HgF both systems



Emission spectrum of HgF $\Delta\lambda = 0$ sequence of 2560 system

(*Facing p* 112)

The three infinite harmonic series and their sums (with topical reference to the Newton and Leibniz series for π)

By F. Soddy, F.R.S.

(*Read* 17 *December* 1943—*Received* 20 *August* 1942—*Revised* 22 *March* 1943)

The infinite harmonic series as hitherto understood, whether with alternating or continuous signs, is not the complete series. It may be extended in both directions to infinity, and then it exists in two forms, either with alternating or with continuous signs, both of which have finite sums.

These are termed ${}^{A}S_D$ and ${}^{C}S_D$ in contradistinction to the singly infinite harmonic series ${}^{A}S_S$—or, since the form with alternating signs alone is convergent, S_S. Examples of both the former and their sums are given in the *Epistola posterior* of Newton to Oldenberg (24 October 1676):

$$ {}^{A}_{4}S^{1}_{D} \quad \text{or} \quad \sum_{M=-\infty}^{M=+\infty} \frac{(-1)^M}{4M+1} = 1+\tfrac{1}{3}-\tfrac{1}{5}-\tfrac{1}{7}+\tfrac{1}{9}+\tfrac{1}{11}--++ \;.\infty = \frac{\pi}{2\sqrt{2}}, $$

$$ {}^{C}_{8}S^{1}_{D} \quad \text{or} \quad \sum_{M=-\infty}^{M=+\infty} \frac{+1}{8M+1} = 1-\tfrac{1}{7}+\tfrac{1}{9}-\tfrac{1}{15}+\tfrac{1}{17}-\tfrac{1}{23}+\tfrac{1}{25}-+ \;.\infty = \frac{\pi(\sqrt{2}+1)}{8}. $$

The Leibniz inverse tangent series (1673), which is the special case of Gregory's general form (1671), for the magnitude of the angle, $\pi/4$, the tangent of which is unity, uniquely may be written not only as ${}_{2}S^{1}_{S}$, which for this case is $\frac{1}{2}[{}^{A}_{2}S^{1}_{D}]$, but also as the complete infinite harmonic series of continuous sign with $a/d = \frac{1}{4}$,

$$ {}^{C}_{4}S^{1}_{D} = \infty \;.\; ++\frac{1}{-11}+\frac{1}{-7}+\frac{1}{-3}+1+\frac{1}{5}+\frac{1}{9}+\frac{1}{13}++ \qquad \infty = \frac{\pi}{4}. $$

The general solution for the sums, in both cases based on a well-known eighteenth-century theorem of Euler, was virtually given by Glaisher (1873) and they may be written

$$ {}^{A}_{d}S^{a}_{D} \quad \text{or} \quad \sum_{-\infty}^{+\infty} \frac{(-1)^M}{a+Md} = \frac{\pi}{d}\operatorname{cosec}\frac{a\pi}{d}, \qquad {}^{C}_{d}S^{a}_{D} \quad \text{or} \quad \sum_{-\infty}^{+\infty} \frac{+1}{a+Md} = \frac{\pi}{d}\cot\frac{a\pi}{d}. $$

The above forms multiplied by a (or what comes to the same thing setting a as unity and using the ratio d/a for the 'common difference') bear the geometrical interpretation that the sum, for the alternating sign series, is the ratio of the length of the arc to that of the chord in a circular sector or segment of included angle $(2\pi a)/d$, and, for the continuous sign series, the same ratio multiplied by cos $(a\pi)/d$. The two series are thus completely self-contained periodic circular functions without any restriction whatever on the magnitude of the angles they numerate.

Glaisher's well-known series for higher powers of π (1873) are all doubly infinite series, as

$$ \left[\frac{\pi}{d}\operatorname{cosec}\frac{a\pi}{d}\right]^2 = \sum_{-\infty}^{+\infty}\frac{1}{(a+Md)^2}; \qquad \left[\frac{\pi}{d}\operatorname{cosec}\frac{a\pi}{d}\right]^2\left[\frac{\pi}{d}\cot\frac{a\pi}{d}\right] = \sum_{-\infty}^{+\infty}\frac{1}{(a+Md)^3}; $$

the first showing in this form that for the alternating sign series the sum of the squares of the individual terms is equal to the square of their sum.

From this viewpoint the ordinary singly infinite harmonic series S_S consists of the positive terms of a cotangent series and the negative terms of the corresponding tangent series. Nevertheless, its sum can be found by modern methods using complex numbers when a and d are integers and a/d is not greater than unity.

A table of sums for all cases up to $d = 16$ is included. The sums are of the form of the sum or difference of $\frac{1}{2}[{}^{A}_{d}S^{a}_{D}]$ and a composite logarithmic quantity termed ${}_{d}l^{a}$.

$$ {}_{d}S^{a} = \frac{\pi}{2d}\operatorname{cosec}\frac{a\pi}{d} + {}_{d}l^{a}, \quad {}_{d}S^{d-a} = \frac{\pi}{2d}\operatorname{cosec}\frac{a\pi}{d} - {}_{d}l^{a}, $$

$$ (-1)^{a}\,{}_{d_0}l^{a} = \frac{2}{d}\sum_{n=1}^{n=\frac{d-1}{2}} \cos\frac{a\pi n}{d}\log_{\epsilon}\sec\frac{\pi n}{d}. \qquad (1) $$

$$ {}_{d_0}l^{a_0} = \frac{2}{d}\sum_{n=1}^{n=d/4} \cos\frac{a\pi n}{d}\log_{\epsilon}\cot\frac{\pi n}{2d}. \qquad (2) $$

(1) applies when d is odd, (2) when it is even, n being $2n-1$. If $d/2$ is even, the upper limit of summation in (2) is $(d-2)/4$, not $d/4$

In (1) $\log\sec(\pi n/d)$ is $2\coth^{-1}\cot^{2}(\pi n/2d)$. In (2) $\log\cot(\pi n/2d)$ is $\tanh^{-1}\cos(\pi n/d)$.

The three infinite harmonic series

It has hitherto been somewhat absolutely laid down in all the algebras that the harmonic progression series cannot be summed. Thus to quote Chrystal (1886): 'Notwithstanding the comparative simplicity of the law of its formation, the harmonic series does not belong to the category of series that can be summed.' True, the context—which 'leaves it as a good exercise in algebraic logic to the student to prove that the sum of a harmonic series to n terms cannot be represented by any rational algebraic function of n'—suggests that no more may have been meant than that it differed from the geometric and arithmetic progression series in that there could be no general solution for the sum of any finite number of terms. But in this, surely, it is the arithmetic and geometric progressions that are somewhat exceptional, since few of the infinite series that can be summed can be summed for any finite number of terms Intuitively it would, perhaps, appear that this must be true of any such series that sums, as the complete harmonic series are shown in this communication to sum, to a transcendental number. Nevertheless, as also is shown here, one of the complete infinite harmonic series is divisible into two parts which can be separately summed.

Again, it may be held that the algebras for some reason confine the term, harmonic series, to the one form with continuous sign, for which what has hitherto been regarded as the infinite series is of necessity divergent and has no limiting sum. But as the term, geometric progression series, is equally applied in all the algebras to series with alternating and with continuous signs, there seems to be no logical reason for this limitation. Even so, as regards the complete infinite series of continuous sign, the statement is not true. For, in the harmonic series, the paradox seems before to have escaped notice that though, as previously understood, the series is divergent, that series is not the complete series, but only one-half of it. When extended to infinity in *both* directions, whether the signs are continuous or alternating, the series are, like the singly-infinite harmonic series of alternating sign, of necessity convergent. There are thus three different infinite harmonic series with finite sums, all of which can be summed.

In this communication, general expressions are given for the sums of all three—the two doubly-infinite series being symbolized as ${}^{A}S_{D}$ and ${}^{C}S_{D}$ for the series of alternating and of continuous sign, respectively, and the singly-infinite series of alternating sign as ${}^{A}S_{S}$, or, since this alone has a finite sum, simply S_{S}. These symbols are used both to identify the series as well as to denote their sums, as the context may require The harmonic series is taken in this paper to be defined as a series of any number of terms which are the reciprocals of numbers in arithmetic progression, whether continuous or alternating in sign. In this symbolism, the 'first term' a and the 'common difference' d of the cognate arithmetical progression are denoted, respectively, by a post-superscript and pre-subscript, thus—${}^{A}_{d}S^{a}_{D}$, ${}^{C}_{d}S^{a}_{D}$ and ${}_{d}S^{a}_{S}$. The name, harmonic, derives from acoustics, the successive terms both of the Leibniz series for $\pi/4$, and that of Newton for $\pi/(2\sqrt{2})$ to be cited $-1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \frac{1}{9}, \frac{1}{11}$, etc.—being numerically in the ratio of the wave-lengths of the fundamental tone and the successive harmonics of the closed or 'stopped' organ pipe Originally, Professor Turnbull recalls, they were also known as 'musical series'. In the case of the geometric progression series, whether the signs are continuous or alternating depends on the sign of the common ratio, whereas for ${}^{C}S_{D}$ the sign of the term depends on which side of the zero it is But for ${}^{A}S_{D}$ the alternation of sign is inherent in the nature of the series. As regards S_{S}, it may be regarded either as the positive half only of ${}^{A}S_{D}$, or as a composite of the positive half of one with the negative half of another, and very closely related, ${}^{C}S_{D}$, and the possibility of its general summation is, therefore, as has been indicated, a little surprising

The series of Newton and Leibniz

Professor Plummer, who had made translations of some of Newton's earlier mathematical writings, in readiness for the tercentenary of Newton's birth on 25 December 1942, has most generously made his unpublished MSS. available to me in preparing this paper. A cordial measure of general appreciation is due to him for having made possible an interesting historical discovery just in time for this tercentenary. Newton's best-known series for the evaluation of π—very much the best till then discovered, and given in all the text-books usually under the term, the sine series—occurs in the *Epistola prior*, to Henry Oldenberg, Secretary of the Royal Society, 13 June 1676, for communication to G. W. Leibniz. But in the *Epistola posterior*, dated 24 October 1676, there is another series, also fairly well known as Newton's, though not always correctly reproduced, very intriguing in the present connexion as it appears to have been the first unequivocal example of a doubly-infinite harmonic series of alternating sign. But, in looking up this, another was found in the same letter, not before specially associated with the name of Newton, which turns out to be the first unequivocal example of the doubly-infinite series of continuous sign also. The first, as Newton wrote it, and as it has been written (when correctly written) for over 265 years, is

$$'1+\tfrac{1}{3}-\tfrac{1}{5}-\tfrac{1}{7}+\tfrac{1}{9}+\tfrac{1}{11}-\tfrac{1}{13}-\tfrac{1}{15}, \text{ etc},$$

for the length of the quadrantal arc of which the chord is unity'—that is, for $\pi/(2\sqrt{2})$. When written as ${}^{A}_{a}S^{a}_{D}$ (since in general ${}^{A}_{a}S^{a}_{D}$ is the same as ${}^{A}_{a}S^{a-a}_{D}$), it is either

$${}^{A}_{4}S^{1}_{D} \quad \text{or} \quad \sum_{M=-\infty}^{M=+\infty} \frac{(-1)^M}{4M+1} = \infty \ldots - + - - \frac{1}{-11} + \frac{1}{-7} - \frac{1}{-3} + 1 - \frac{1}{5} + \frac{1}{9} - + - \ldots \infty = \frac{\pi}{2\sqrt{2}},$$

or

$${}^{A}_{4}S^{3}_{D} \quad \text{or} \quad \sum_{M=-\infty}^{M=+\infty} \frac{(-1)^M}{4M+3} = \infty \ldots - + - \frac{1}{-9} + \frac{1}{-5} - \frac{1}{-1} + \frac{1}{3} - \frac{1}{7} + \frac{1}{11} - + - \infty = \frac{\pi}{2\sqrt{2}}.$$

The second series, the prototype of ${}^{C}_{a}S^{a}_{D}$, which Professor Plummer pointed out is obtained from the first by addition to the Leibniz series for $\pi/4$ and dividing by 2, occurs as a device for shortening the computation of π by either of these series:

'For if one wished by the simple calculation of the series $1 + \frac{1}{3} - \frac{1}{5} - \frac{1}{7} + \frac{1}{9}$, etc., to find the length of the quadrant to twenty decimal places, it would need about 5 000 000 000 terms of the series for the calculation of which a thousand years would be required. Through the tangent of 45° the result would be obtained even later. But by using the sine of 45°, fifty-five or sixty places of this series

$$\sqrt{\tfrac{1}{2}} \times (1 + \tfrac{1}{12} + \tfrac{3}{160} + \tfrac{5}{896}, \text{etc.})$$

would suffice; the computation of which in my opinion could be finished in three days or four....

'By the series of Leibniz (i.e. Gregory), if half the term in the last place be added and some other like devices be employed, the computation can be carried to many decimals. As by assuming the sum of the terms

$$1 - \tfrac{1}{7} + \tfrac{1}{9} - \tfrac{1}{15} + \tfrac{1}{17} - \tfrac{1}{23} + \tfrac{1}{25} - \tfrac{1}{31} + \tfrac{1}{33}, \text{ etc},$$

is to the whole series

$$1 - \tfrac{1}{3} + \tfrac{1}{5} - \tfrac{1}{7} + \tfrac{1}{9} - \tfrac{1}{11} +, \text{ etc},$$

as $1+\sqrt{2}$ to 2.'

This second series is

$${}^{C}_{8}S^{1}_{D} \quad \text{or} \quad \sum_{M=-\infty}^{M=+\infty} \frac{+1}{8M+1}$$

$$= \infty \ldots + + \frac{1}{-23} + \frac{1}{-15} + \frac{1}{-7} + 1 + \frac{1}{9} + \frac{1}{17} + \frac{1}{25} + + \ldots \infty = \frac{\pi(\sqrt{2}+1)}{8}.$$

The Leibniz series (1673) which is usually described, quite inadequately, as 'merely' a special case of Gregory's inverse tangent series (1671)

$$\tan^{-1} u = u - \tfrac{1}{3}u^3 + \tfrac{1}{5}u^5 - \tfrac{1}{7}u^7 + \tfrac{1}{9}u^9 - + - \ldots \infty,$$

for the magnitude of the angle, $\pi/4$ or 45°, for which the tangent, u, is unity, is quite unique. For not only is it the exact half of ${}^{A}_{2}S^{1}_{D}$,

$${}^{A}_{2}S^{1}_{D} = \infty \ldots + - \frac{1}{-9} + \frac{1}{-7} - \frac{1}{-5} + \frac{1}{-3} - \frac{1}{-1} + 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - + \ldots \infty = \frac{\pi}{2},$$

but also it is the complete ${}^{C}_{4}S^{1}_{D}$ continuous-sign series, with a, 1 and d, 4—

$${}^{C}_{4}S^{1}_{D} \quad \text{or} \quad \sum_{+\infty}^{-\infty} \frac{+1}{4M+1} = \infty \ldots + + \frac{1}{-11} + \frac{1}{-7} + \frac{1}{-3} + 1 + \frac{1}{5} + \frac{1}{9} + \frac{1}{13} + + \ldots \infty = \frac{\pi}{4}.$$

The double paradox is that ${}^{A}_{d}S^{a}_{D}$ appears when written in numerical sequence as the sum of two different, but related, singly-infinite harmonic series—${}^{A}_{d}S^{a}_{D} = {}^{A}_{d}S^{a}_{S} + {}^{A}_{d}S^{d-a}_{S}$—whereas the *continuous* sign series, ${}^{C}_{d}S^{a}_{D}$, appears to have regularly *alternating* signs, but, except for the Leibniz series, *not* to be numerically harmonic. It has to be represented, for example, as the difference of two related singly-infinite series of continuous sign—${}^{C}_{d}S^{a}_{D} = {}^{C}_{d}S^{a}_{S} - {}^{C}_{d}S^{d-a}_{S}$—and is an instance of how a convergent series may result from the difference of two related divergent series. Thus Newton was the first to give unequivocal examples of *both* of the doubly-infinite harmonic series, the Gregory series, which preceded these by three years, being the special case where all three types of infinite harmonic series with finite sums are the same or similar.

The case where a and d are both unity—${}^{A}_{1}S^{1}_{S} = \log_{e} 2$—in doubly-infinite form, ${}^{A}_{1}S^{0}_{D}$ and ${}^{C}_{1}S^{0}_{D}$, is also unique in that the sum in each case is infinite by virtue of the single central term, $1/0$, the rest of the terms cancelling one another in two different ways,

$${}^{A}_{1}S^{0}_{D} = -\log_{e} 2 + \tfrac{1}{0} + \log_{e} 2, \quad {}^{C}_{1}S^{0}_{D} = -{}^{C}_{1}S^{1}_{S} + \tfrac{1}{0} + {}^{C}_{1}S^{1}_{S},$$

those on either side being numerically identical but of opposite sign. But these series are not in any proper sense of the word divergent. On the contrary, the sum is exactly the same, however many terms be taken, so long as they are symmetrical with reference to the $M = 0$, or zero, term.

The sums of the two doubly-infinite harmonic series

The general solutions for the sums of both types of doubly-infinite harmonic series are given by the eighteenth-century theorem of Euler, known sometimes as the expansion of the cotangent in partial fractions, though hitherto, naturally, this can hardly be expected to have been recognized, since the series themselves seem to have escaped recognition. But they were virtually given by Glaisher (1873), as series without number by which π can be evaluated, in the forms

$$\pi = n \sin\frac{\pi}{n}\left[1 + \frac{1}{n-1} - \frac{1}{n+1} - \frac{1}{2n-1} + \frac{1}{2n+1} + - - + + \ldots \infty\right], \tag{1}$$

$$\pi = n \tan\frac{\pi}{n}\left[1 - \frac{1}{n-1} + \frac{1}{n+1} - \frac{1}{2n-1} + \frac{1}{2n+1} - + - + \ldots \infty\right], \tag{2}$$

the second being the Euler cotangent theorem, and the first being simply derivable from it by the relation $\operatorname{cosec}\theta = \cot\theta/2 - \cot\theta$, which, incidentally, is that connecting the two prototypes due to Newton. It suffices merely to replace n in

these expressions by d/a to obtain the general solutions for the sums of both types.

$$\sum_{-\infty}^{+\infty} \frac{(-1)^M}{M\frac{d}{a}+1} = \frac{a\pi}{d}\operatorname{cosec}\frac{a\pi}{d}; \tag{3}$$

$$\sum_{-\infty}^{+\infty} \frac{+1}{M\frac{d}{a}+1} = \frac{a\pi}{d}\cot\frac{a\pi}{d}. \tag{4}$$

Dividing these by a gives the solutions for the sum of the infinite harmonic series, as here defined:

$${}^{A}_{d}S^{a}_{D} = \sum_{-\infty}^{+\infty} \frac{(-1)^M}{a+Md} = \frac{\pi}{d}\operatorname{cosec}\frac{a\pi}{d}; \tag{5}$$

$${}^{C}_{d}S^{a}_{D} = \sum_{-\infty}^{+\infty} \frac{+1}{a+Md} = \frac{\pi}{d}\cot\frac{a\pi}{d}. \tag{6}$$

These expressions show that these series (multiplied by a if a is not unity) are simple completely self-contained periodic circular functions, without any restriction whatever on the magnitude of the angles which they numerate, though, as for the other trigonometrical ratios, in their geometrical interpretation appropriate sign conventions must be adopted for 'angles' greater than π for the first and $\pi/2$ for the second. The first, $a \times {}^{A}_{d}S^{a}_{D}$, is in fact, the trigonometrical ratio earliest to be used, by the Greeks. It is the ratio of the length of the arc to that of the chord in a circular sector or segment of included angle $(2a\pi/d)$. The terms in which Newton stated the sum of his series for which a/d is 1/4, 'the length of the quadrantal arc for which the chord is unity' were, prophetically, quite general, if 'quadrantal arc' is generalized to the (a/d)th part of the circle That for Gregory's series, for example, for which a/d is 1/2, doubled by extension to infinity backwards, sums to $\pi/2$, the arc/chord ratio of the semicircular arc. The second, $a \times {}^{C}_{d}S^{a}_{D}$, is the same arc/chord ratio, multiplied by that of the apothem to the radius, as the Greeks would have said, or $\cos(a\pi/d)$. Thus for the Gregory series summing to $\pi/4$, it is $\pi/(2\sqrt{2}) \times 1/\sqrt{2}$, a/d being 1/4.

The corresponding forms for the secant and tangent, using a' to denote $d/2 \pm a$, are

$${}^{A}_{d}S^{a'}_{D} \quad \text{or} \quad \sum_{-\infty}^{+\infty} \frac{(-1)^M}{Md+d/2\pm a} = \frac{\pi}{d}\sec\frac{a\pi}{d}; \tag{7}$$

$${}^{C}_{d}S^{a'}_{D} \quad \text{or} \quad \sum_{-\infty}^{+\infty} \frac{1}{Md+d/2\pm a} = \mp\frac{\pi}{d}\tan\frac{a\pi}{d}. \tag{8}$$

Thus, subtracting from Newton's series for $\pi/(2\sqrt{2})$ that of Leibniz for $\pi/4$, and dividing by 2, instead of adding the series, gives ${}^{C}_{8}S^{3}_{D}$, instead of ${}^{C}_{8}S^{1}_{D}$, which sums to

$$\frac{\pi}{8}\cot\frac{3\pi}{8} = \frac{\pi}{8}\tan\frac{\pi}{8} = \frac{\pi(\sqrt{2}-1)}{8},$$

in conformity with $\cot(\pi/2-\theta) = \tan\theta$ and with $\operatorname{cosec}\theta - \cot\theta = \tan\theta/2$. The relation, derived from this last by combining with it the corresponding addition relation, $\operatorname{cosec}\theta + \cot\theta = \cot\theta/2$,

$$\begin{aligned}\operatorname{cosec}\theta = \tfrac{1}{2}(\cot\theta/2 + \tan\theta/2) &= \tfrac{1}{2}\{\cot\theta/2 + \cot(\pi/2-\theta)\}\\ &= \tfrac{1}{2}\{\cot\theta/2 - \cot(\pi/2+\theta)\},\end{aligned}$$

represents the splitting of ${}^{A}_{d}S^{a}_{D}$ into ${}^{C}_{2d}S^{a}_{D} - {}^{C}_{2d}S^{d+a}_{D}$, by taking alternate terms

$${}^{A}_{d}S^{a}_{D} = \left\{\begin{array}{l|l} \cdots + \dfrac{1}{a-4d} + \dfrac{1}{a-2d} & + \dfrac{1}{a} + \dfrac{1}{a+2d} + \dfrac{1}{a+4d} + \cdots = {}^{C}_{2d}S^{a}_{D} \\ \cdots - \dfrac{1}{a-3d} - \dfrac{1}{a-d} & \underbrace{- \dfrac{1}{a+d} - \dfrac{1}{a+3d} - \cdots}_{{}^{A}_{d}S^{a}_{S}} = -{}^{C}_{2d}S^{d+a}_{D}. \end{array}\right.$$

The terms to the right of the dividing line constitute ${}_{d}S^{a}_{S}$, which is thus simply shown to consist of the positive terms of a cotangent series and the negative terms of the corresponding tangent series, both with d doubled.

Extending the series in both directions to infinity, by bringing in every positive and negative value of every natural number once, makes the sign of M and that of d/a on the left in expressions (3) and (4) immaterial. On the left the series is unchanged and on the right the sign of d/a is changed twice if at all, and, since

$$\operatorname{cosec}(-\theta) = -\operatorname{cosec}\theta \quad \text{and} \quad \cot(-\theta) = -\cot\theta,$$

the product is unchanged. It is only the range of d/a from unity to zero that enters into the composition of the series. Integral changes of d/a change $(d/a)\,M$ to, say, $(d/a+n)\,M = d/a\,M + nM$, that is, they are the same as integral changes of M, which, since all values of M are represented in the series once and once only, do not affect the numerical values of the terms at all, or their sign in (4). In (3) if n is odd there is a change of sign by virtue of the $(-1)^{M}$ numerator. In this way the series conform to $\operatorname{cosec}\theta = \operatorname{cosec}\{n\pi + (-1)^{n}\theta\}$ for (3) and to $\cot\theta = \cot(n\pi+\theta)$, where n is any integer. When a/d is integral, the term for which $-M$ has this value becomes $1/0$, the sum becomes infinite by virtue of this single term, the rest of the terms cancelling as already stated, conforming to the infinite value both of cosecant and cotangent for angles which are integral multiples of π. When a/d is $1/2$, the terms for the positive and negative halves of the series are numerically identical, of the same sign for (3) so that the series sums to $\pi/2$, in conformity with $\operatorname{cosec}\pi/2$ being unity, and of opposite sign for (4) conforming to $\cot\pi/2$ being zero. But the feature that justifies the harmonic series being considered, not 'merely' a special case of the power series, as those for the sine and tangent, for which the powers involved are those of unity, but as a class to themselves, is that, in sharp contrast to such series, their convergence remains quite unaffected by the magnitude, however great, of the angles they numerate.

Glaisher's series for the powers of π

Glaisher (1873), by two successive differentiations of his expression (2) with regard to π/n, obtained the following now well-known series for the square and the cube of π:

$$\pi^2 = n^2 \sin^2\left(\frac{\pi}{n}\right)\left[1 + \frac{1}{(n-1)^2} + \frac{1}{(n+1)^2} + \frac{1}{(2n-1)^2} + \frac{1}{(2n+1)^2} + + \ldots \infty\right], \tag{9}$$

$$\pi^3 = n^3 \sin^2\left(\frac{\pi}{n}\right)\tan\frac{\pi}{n}\left[1 - \frac{1}{(n-1)^3} + \frac{1}{(n+1)^3} - \frac{1}{(2n-1)^3} + \frac{1}{(2n+1)^3} - + - \ldots \infty\right]. \tag{10}$$

These also can be expressed more concisely and informatively in doubly-infinite form as

$$\left[\frac{\pi}{d}\operatorname{cosec}\frac{a\pi}{d}\right]^2 = \sum_{-\infty}^{+\infty} \frac{1}{(a+Md)^2}, \tag{11}$$

$$\left[\frac{\pi}{d}\operatorname{cosec}\frac{a\pi}{d}\right]^2\left[\frac{\pi}{d}\cot\frac{a\pi}{d}\right] = \sum_{-\infty}^{+\infty} \frac{1}{(a+Md)^3}. \tag{12}$$

The first shows that the harmonic series which numerates the arc/chord ratio of the circle has the remarkable property that the sum of the squares of the separate terms is equal to the square of their sum This is reminiscent of the relationships governing the curvatures of four mutually kissing circles and of five mutually kissing spheres (1936). The second shows that if ${}^{c}_{d}S^{a}_{D}$ is multiplied by the squared ${}^{1}_{d}S^{a}_{D}$, the separate terms of the former series are cubed. Some similar results are given—(13) is from (10), (14) from (8), (15) from (7) and (16) from (1), all by simple differentiation:

$$\left[\frac{\pi}{d}\right]^4\left[\operatorname{cosec}^2\left(\frac{a\pi}{d}\right)\right]\left[\operatorname{cosec}^2\left(\frac{a\pi}{d}\right) - \frac{2}{3}\right] = \sum_{-\infty}^{+\infty} \frac{1}{(a+Md)^4}; \tag{13}$$

$$\left[\frac{\pi}{d}\sec\frac{a\pi}{d}\right]^2 = \sum_{-\infty}^{+\infty} \frac{1}{(Md+d/2+a)^2}; \tag{14}$$

$$\left(\frac{\pi}{d}\right)^2 \sec\frac{a\pi}{d}\tan\frac{a\pi}{d} = \sum_{-\infty}^{+\infty} \frac{(-1)^M}{(Md-d/2+a)^2}; \tag{15}$$

$$\left(\frac{\pi}{d}\right)^2 \operatorname{cosec}\frac{a\pi}{d}\cot\frac{a\pi}{d} = \sum_{-\infty}^{+\infty}\left[\frac{1}{(2Md+a)^2} - \frac{1}{(2Md+d+a)^2}\right]. \tag{16}$$

(16) could be now more simply derived from $4\operatorname{cosec}\theta \cdot \cot\theta = \operatorname{cosec}^2(\theta/2) - \sec^2(\theta/2)$ (15) and (16) show that alternating sign series can be obtained for an even power of π, the alternation resulting inherently for (15) and as the difference between two continuous sign series for (16) For $a/d = \frac{1}{4}$, when the two are identities, there results the curious series

$$\frac{\sqrt{2}\pi^2}{16} = 1 - \frac{1}{3^2} - \frac{1}{5^2} + \frac{1}{7^2} + \frac{1}{9^2} - \frac{1}{11^2} - \frac{1}{13^2} + + - - \ldots \infty.$$

It is, to say the least, surprising that these doubly-infinite forms, which in the calculus of infinitesimals are as common as the singly infinite, should apparently not have been made use of previously in the calculus of finite differences, though no

special examination of this point has yet been made. There is, however, at least one very interesting example (Hobson 1928*a*) where relation (11) is given as an exercise—'Prove that $\sum_{n=-\infty}^{n=+\infty} \frac{1}{(n\pi+\theta)^2} = \operatorname{cosec}^2\theta$'—and which renders it the more curious that the applicability of the form to the doubly-infinite harmonic series of Newton should, apparently, have escaped notice.

The sum of the singly-infinite harmonic series

The problem—to find a general solution for the sum of the singly-infinite harmonic series, S_S—can now be somewhat narrowed. Since ${}^{A}_{d}S^{a}_{D} = {}_{d}S^{a}_{S} + {}_{d}S^{d-a}_{S} = \frac{\pi}{d}\operatorname{cosec}\frac{a\pi}{d}$, S_S is known if ${}^{A}_{d}S^{a}_{S} - {}^{A}_{d}S^{d-a}_{S}$ can be found. Calling this $2\,{}_{d}l^{a}$, and ${}_{d}l^{a}$ the difference term,

$$ {}_{d}S^{a}_{S} = \frac{\pi}{2d}\operatorname{cosec}\frac{a\pi}{d} + {}_{d}l^{a} \quad \text{and} \quad {}^{A}_{d}S^{d-a}_{S} = \frac{\pi}{2d}\operatorname{cosec}\frac{a\pi}{d} - {}_{d}l^{a}. $$

Many ineffectual attempts were made by simple methods to find S_S or ${}_{d}l^{a}$, but two results were obtained from this enquiry which may first be mentioned as throwing some light on the problem

The first was that the pair ${}_{3}S^{1}_{S}$ and ${}_{4}S^{2}_{S}$ can be summed by inspection on comparing it with ${}^{C}_{3}S^{1}_{D}$:

$$ \begin{aligned} {}_{3}S^{1}_{S} &= \quad 1 - \tfrac{1}{4} + \tfrac{1}{7} - \tfrac{1}{10} + \tfrac{1}{13} - \tfrac{1}{16} + \tfrac{1}{19} - \tfrac{1}{22} + \tfrac{1}{25} - \tfrac{1}{28} + \tfrac{1}{31} - \tfrac{1}{34} + - \ldots, \\ {}^{C}_{3}S^{1}_{D} &= \left.\begin{matrix} 1 + \tfrac{1}{2} + \tfrac{1}{7} + \tfrac{1}{10} + \tfrac{1}{13} + \tfrac{1}{16} + \tfrac{1}{19} + \tfrac{1}{22} + \tfrac{1}{25} + \tfrac{1}{28} + \tfrac{1}{31} + \tfrac{1}{34} + + \ldots \\ -\tfrac{1}{2} - \tfrac{1}{5} - \tfrac{1}{8} - \tfrac{1}{11} - \tfrac{1}{14} - \tfrac{1}{17} - \tfrac{1}{20} - \tfrac{1}{23} - \tfrac{1}{26} - \tfrac{1}{29} - \tfrac{1}{32} - \tfrac{1}{35} - - \quad . \end{matrix}\right\} = \frac{\pi}{3\sqrt{3}}, \\ {}_{3}S^{2}_{S} &= \quad \tfrac{1}{2} - \tfrac{1}{5} + \tfrac{1}{8} - \tfrac{1}{11} + \tfrac{1}{14} - \tfrac{1}{17} + \tfrac{1}{20} - \tfrac{1}{23} + \tfrac{1}{26} - \tfrac{1}{29} + \tfrac{1}{32} - \tfrac{1}{35} + - \ldots \end{aligned} $$

Comparing first ${}_{3}S^{1}_{S}$ and ${}^{C}_{3}S^{1}_{D}$, and removing like terms, all the remaining terms of ${}_{3}S^{1}_{S}$ can be removed by a pair in ${}^{C}_{3}S^{1}_{D}$, one in the upper and one in the lower line, as $-\frac{1}{4}$ with $(\frac{1}{2}-\frac{1}{3})$, $-\frac{1}{10}$ with $(\frac{1}{10}-\frac{1}{5})$ and so on, leaving of ${}^{C}_{3}S^{1}_{D}$, when an indefinitely large number, N, of terms of ${}_{3}S^{1}_{S}$ are operated upon,

$$ -\sum_{M=N/2}^{M=\infty} \frac{1}{3M+1} = -\frac{1}{3}\sum_{M=N/2}^{M=N} \frac{1}{M} = -\frac{1}{3}\sum_{M=1}^{M=\infty} \frac{(-1)^M}{M} = -\tfrac{1}{3}\log_e 2. $$

Similarly for ${}_{3}S^{2}_{S}$ the remaining terms of ${}^{C}_{3}S^{1}_{D}$ sum to $+\frac{1}{3}\log_e 2$. Hence

$$ {}_{3}S^{1}_{S} = \frac{\pi}{3\sqrt{3}} + \tfrac{1}{3}\log_e 2 \quad \text{and} \quad {}_{3}S^{2}_{S} = \frac{\pi}{3\sqrt{3}} - \tfrac{1}{3}\log_e 2. $$

When the sums are known the result is 'obvious' from the general relation, when d is odd,

$$ {}_{d}S^{1}_{S} - {}_{d}S^{2}_{S} + {}_{d}S^{3}_{S} - + - \ldots + {}_{d}S^{d-2}_{S} - {}_{d}S^{d-1}_{S} = \frac{d-1}{d}\log_e 2, $$

which, for $d = 3$, gives ${}_{3}S^{1}_{S} - {}_{3}S^{2}_{S} = \frac{2}{3}\log_e 2$. But for $d = 5$, of the four series to be summed, although every other combination by pairs can be obtained by similar

simple devices, the ones wanted, ${}_d l^a$, could not, as there always seems to be one more unknown than equations to solve them. For $d = 3$, the solution is possible, because $\operatorname{cosec} \pi/3 = 2 \cot \pi/3$.

The second attempt is more involved and has led as yet to no useful result, but may be worth noting. If, for example, the two series composing ${}_d l^a$ have denominators differing by 2, as is the case for $a = 1$, then replacing them by others in which for each separate term, $1/(p+1)$ and $1/(p-1)$ respectively, the corresponding infinite geometrical progression series, $1-p+p^2-\ldots$ and $1+p+p^2+\ldots$, the arc/chord ratio and ${}_d l^a$ assume similar forms

$$ {}_d^A S_D^1 = {}_d^A S_S^1 + {}_d^A S_S^{d-1} = \frac{\pi}{d} \operatorname{cosec} \frac{\pi}{d} = 1 + 2 \sum_{m=1}^{m=\infty} \sum_{M=1}^{M=\infty} \frac{(-1)^M}{(Md)^{2m}} $$

$$ {}_d^A S_S^1 - {}_d^A S_S^{d-1} = 2 \, {}_d l' = 1 + 2 \sum_{m=1}^{m=\infty} \sum_{M=1}^{M=\infty} \frac{(-1)^M}{(Md)^{2m+1}}. $$

In this example, the former has the even and the latter the odd powers of m in the terms, but this may or may not be significant as in other cases, for example, in operating similarly on the ${}_{2d} S_S^{d-1} \pm {}_{2d} S_S^{d+1}$ pair, the converse is the case. This seems to establish that ${}_d l^a$ is a quantity similar in nature to the arc/chord ratio but its geometrical interpretation has not yet been achieved.

The series S_S can be summed by the use of complex numbers in a form satisfying the modern sense of the term, summation, that is, the sum can be expressed by a number of terms having no relation to the number of terms in the series, except when the d/a ratio is itself transcendental, for which the one infinite series would be replaced by another. In other words d and a must be integers, when the number of terms in the solution is a function of d only. Also, so far as yet examined, it would appear that a/d must not be greater than unity, that is, the summation of any finite number of terms of S_S is still not possible, as Chrystal in 1886 enjoined his students to prove for themselves for the continuous sign series.

In the current text-books on infinite series and on trigonometry, among fairly numerous examples of what have been shown to be doubly-infinite harmonic series, summed by Euler's theorem, the sum was found for three of the simpler cases of S_S—the above pair with $d = 3$, ${}_4 S_S^1$ and ${}_6 S_S^1$—but hitherto it seems to have been regarded 'merely' as a special case of more general forms and no general solution attempted or obtained. The clue was given by the last, the sum of which is given by Hobson (1928*b*) as an exercise in the form: 'Prove that

$$ \frac{\tan^{-1}\alpha}{\alpha} + \frac{\tan^{-1}\beta}{\beta} + \frac{\tan^{-1}\gamma}{\gamma} = \frac{\pi}{2} + \frac{\sqrt{3}}{4} \log \frac{2+\sqrt{3}}{2-\sqrt{3}} = 3[1 - \tfrac{1}{7} + \tfrac{1}{13} - \tfrac{1}{19} + \tfrac{1}{25} - \ldots], $$

where α, β, γ are the three cube-roots of unity.' This example well illustrates the dual character of the problem, which may be stated as follows: To find, first, a particular combination of the roots of unity which, expanded as the power terms of a series containing the terms of the required S_S, as the factors, sum to zero in

all the terms of the series, except those in S_S, and to the latter assigns the real root, $+1$ if the number is odd, or roots, $+1$ and -1 if the number is even, with the correct sign. This part of the problem being arbitrary, the rules to be observed are likewise arbitrary. The second part is to transform the particular combination found effective into real quantities only, which, the problem being essentially geometrical, is always possible.

In the above example, only odd numbers enter into the terms of the series, ${}_6S^1_S$, and Gregory's inverse tangent series satisfies the conditions. Out of four possible combinations of odd and even values of a and d—reduced in practice to three, by virtue of the relation ${}_{pd}S^{ap}_S = \frac{1}{p}[{}_dS^a_S]$—it can be used for this one only. The rules are (1) to use, as above, the $d/2$ $(d/2)$th roots of unity when $d/2$ is odd, and (2) when $d/2$ is even, the first half, $d/2$ in number, of the odd-power $(2d)$th roots. Thus the solution by this method for ${}_4S^1_S$ is

$${}_8\omega^{-1}\tan^{-1}{}_8\omega^1 + {}_8\omega^{-3}\tan^{-1}{}_8\omega^3 = 2[1 - \tfrac{1}{5} + \tfrac{1}{9} - {}_1{}_3^1 + - + \ldots\infty],$$

$${}^1_4S^1_S = \frac{\pi}{4\sqrt{2}} + \frac{1}{2\sqrt{2}}\log_e(\sqrt{2}+1)$$

The other series available

$$\log_e(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 + - + \ldots\infty$$

has in the denominators all the natural numbers, even and odd, and, with a corresponding pair of rules, serves to sum any S_S with integral values of a and d, a/d not exceeding unity. These rules are: (1) d odd, or d_o, use the d (d)th roots; (2) d even, or d_e, use the d odd-power $(2d)$th roots.

Further, in both cases, the a of the series appears as a power of the root by which the term is divided, the general form of the single term of the combination being $\omega^{-an}\tan^{-1}\omega^n$ and $\omega^{-an}\log_e(1+\omega^n)$ respectively. It is remarkable how simply this (pre-Eureka!) formidable complication in generalizing the above example is accommodated.

The general solutions in the two cases can then be expressed in the first instance as

$$\left(\frac{d_e}{2}\text{odd}\right)\quad {}_{d_e}S^{a_e}_S = \frac{2}{d}\sum_{n=1}^{n=d/2} {}_{d/2}\omega^{-an}\tan^{-1}{}_{d/2}\omega^n;$$

$$\left(\frac{d_e}{2}\text{even}\right)\quad {}_{d_e}S^{a_e}_S = \frac{1}{d}\sum_{n=1}^{n=d/2} {}_{2d}\omega^{-am}\tan^{-1}{}_{2d}\omega^m;$$

$${}_{d_o}S^{a_o}_S = \frac{(-1)^{a+1}}{d}\sum_{n=1}^{n=d} {}_d\omega^{-an}\log_e(1+{}_d\omega^n);$$

$${}_{d_e}S^{a_e}_S = \frac{1}{d}\sum_{n=1}^{n=d} {}_{2d}\omega^{-am}\log_e(1+{}_{2d}\omega^m);$$

where $${}_d\omega^n = \cos\frac{2\pi n}{d} + i\sin\frac{2\pi n}{d} \quad \text{and} \quad m = 2n-1$$

In the second step, the reduction to real forms, the requisite formulae are

$$\tan^{-1}(\pm\cos\theta+i\sin\theta)=\pm\frac{\pi}{4}+\tfrac{1}{2}i\log_e\tan\left(\frac{\pi}{4}+\frac{\theta}{2}\right);$$

$$\tan^{-1}(\pm\cos\theta-i\sin\theta)=\pm\frac{\pi}{4}-\tfrac{1}{2}i\log_e\tan\left(\frac{\pi}{4}+\frac{\theta}{2}\right);$$

$$\log_e(1+\cos\theta\pm i\sin\theta)=\tfrac{1}{2}\log_e(2+2\cos\theta)\pm i\frac{\theta}{2};$$

$$\log_e(1-\cos\theta\pm\sin\theta)=\tfrac{1}{2}\log_e(2-2\cos\theta)\pm i\left(\frac{\pi}{2}-\frac{\theta}{2}\right).$$

(The + or − sign on the right conforms to that on the left in each of the four formulae.) In practice it is better to operate with the difference pair, rather than with the separate series, which eliminates the imaginaries at once, and then to add ${}_dl^a$ to, or subtract it from, $\frac{\pi}{2d}\operatorname{cosec}\frac{a\pi}{d}$ to get ${}_dS_S^a$ or ${}_dS_S^{d-a}$ respectively. The two series yield identities when either may be used. In their final form the general solutions are [(17) applying when d is odd, (18) when d is even, $m=2n-1$, and when $d/2$ is even, the upper limit of summation is $(d-2)/4$, not $d/4$]—

$$(-1)^a{}_{d_0}l^a=\frac{2}{d}\sum_{n=1}^{n=(d-1)/2}\cos\frac{2a\pi n}{d}\log_e\sec\frac{\pi n}{d}; \tag{17}$$

$${}_{d_e}l^{a_0}=\frac{2}{d}\sum_{n=1}^{n=d/4}\cos\frac{a\pi m}{d}\log_e\cot\frac{\pi m}{2d}. \tag{18}$$

In (17) for $\log\sec(\pi n/d)$, $2\coth^{-1}\cot^2(\pi n/2d)$ and in (18) for $\log\cot(\pi m/2d)$, $\tanh^{-1}\cos(\pi m/d)$ may be substituted, so that for the simpler forms ${}_dl^a$ consists of a single term, the product of a circular function and inverse hyperbolic function. Further, in connexion with the geometric interpretation of ${}_dl^a$, it may be remarked, that when, instead of ${}_dl^a$, the single S_S is calculated, the part $(\pi/2d)\operatorname{cosec}(a\pi/d)$ appears as a series of sines of multiple angles, given by (19) and (20), which, but for the considerations with which this section began, might not have been recognized as cosecants.

$$\operatorname{cosec}\frac{a\pi}{d_e}=\frac{2}{d}\sum_{n=1}^{n=d/2}m\sin\frac{a\pi m}{d}; \tag{19}$$

$$\operatorname{cosec}\frac{a\pi}{d_0}=(-1)^{a+1}\frac{4}{d}\sum_{n=1}^{n=\frac{d-1}{2}}n\sin\frac{2a\pi n}{d}. \tag{20}$$

(19) for d even reduces readily, by combination of terms equidistant from the ends, to the ordinary expression for a series of cosines of angles in arithmetic progression, but (20) for d odd is less tractable.

Thus the singly-infinite harmonic series, itself part only of a doubly-infinite series summing to a transcendental number, can also be generally summed. That

previously this appeared to the author to be very unlikely may have arisen in part from the extension of the meaning of the term 'summation' of recent years to series which sum to transcendental quantities. For, a century ago, possibly even when Chrystal wrote, neither series would have been considered really 'summed'. The method employs the various imaginary roots of unity to modify series, of which (so far at least) there are only two known as effective, each of which starts from unity and itself sums to a known transcendental number. This does not in itself seem to imply any general possibility of partial summation of series summing to transcendental numbers but rather that the harmonic series may be exceptional in so far as two singly infinite forms exist by which the transcendental numbers π and e are, or can be, expressed in terms of the natural numbers.

Table of sums

A table accompanies this paper giving the composition and numerical value of ${}_dl^a$ and the numerical values of ${}_dS_S^a$ and ${}_dS_S^{d-a}$ for all the necessary values up to $d = 16$, only those with a prime to and less than one-half d being independent series. The numerical values of ${}_dl^a$ and ${}_dS^{d-a}$ have been calculated from and have the accuracy of Chambers *Mathematical Tables*, the values being given to seven places of decimals with the last uncertain Up to this accuracy, which involves that of the cosecant as well as of ${}_dl^a$, they have been checked by entirely independent computation of ${}_dS_S^a$ by the method recently published (1942). But these computations are taken to from 12 to 15 places, the last being accurate The computations being made entirely without mechanical aid, the best method was found to be to do successive computations of each series with increasing values for n till the target accuracy was attained, as in this way errors could be more readily detected. It is quite idle to compute more than a very few 'small terms' by this method, five to ten at most, or to proceed further after the convergence ratio rises above 0 2 to 0·25, owing to the impracticability of estimating accurately the unincluded remainder The latter must be made relatively negligible Values of n from 10 to 20 were used for the final computations. The main quantity was computed by the method of 'summate numerators' each term of the $n-1$ terms being computed by long division and checked at once by remultiplication To determine from these computations ${}_dS_S^{d-a}$ to the same accuracy would naturally require the cosecant to be known to the same number of places.

Incidentally, this computation method, for ${}_dA_D^a$ of course not for ${}_dS_S^a$, would now probably be one of the quickest and most powerful ways of computing the cosecant and other trigonometrical ratios *ab initio*. For the cosecant of an angle D^0, a being $D/180$, the formula is

$$\operatorname{cosec} D^0 = \frac{n!}{2^{n-1}\pi} \sum_{M=-\infty}^{M=+\infty} \frac{1}{(2M+a)(2M+1+a)(2M+2+a)\dots(2M+n+a)}$$

$(n+1)$ factors $\qquad (n = 1, 2, 3, \dots, \infty).$

SUMS OF THE SINGLY-INFINITE HARMONIC SERIES. $\frac{A}{d}S^a_S = \frac{1}{a} - \frac{1}{a+d} + \frac{1}{a+2d} - \ldots \infty$ [UP TO $d = 16$]

d	a	$d^a - \frac{2}{d}$	d^a	$\frac{A}{d}S^a_S$	$\frac{A}{d}S^{a-1}_S$
1	1	—	—	$\int_1^1$ 0 6931 4718 0599 453 $= \log_e 2$	
2	1	—	—	$\int_2^1$ 0 7853 9816 3397 448 $= \pi/4$	
3	1	$+1I$	0·231 0491	$\int_3^1$ 0 8356 4884 8264 75	0·373 5507 $\int_3^2$
4	1	$+11$	0 311 6126	$\int_4^1$ 0 8669 7298 7339 91	0 243 7478 $\int_4^3$
5	1	$-2I+12$	0 353 8338	$\int_5^1$ 0 8883 1357 2651 789	0 180 6460 $\int_5^4$
	2	$-1I+22$	0 076 5750	$\int_5^2$ 0 4069 0163 4289 45	0 253 7516 $\int_5^3$
6	1	$+11$	0 380 1730	$\int_6^1$ 0·9037 7177 3748 62	0 143 4258 $\int_6^5$
7	1	$-2I+32+13$	0 398 2911	$\int_7^1$ 0 9154 7952 6837 606	0·118 8969 $\int_7^6$
	2	$-3I-12+23$	0 139 4559	$\int_7^2$ 0 4264 7358 3054 084	0 147 5619 $\int_7^5$
	3	$+1I-22+33$	0 038 2275	$\int_7^3$ 0 2683 9778 5359 274	0 191 9429 $\int_7^4$
8	1	$+11+33$	0 411 5656	$\int_8^1$ 0 9246 5170 5775 59	0 101 5204 $\int_8^7$
	3	$+31-13$	0 061 3712	$\int_8^3$ 0·2738 9821 9208 27	0 151 1560 $\int_8^5$
9	1	$-2I-42+33+14$	0 421 7303	$\int_9^1$ 0 9320 3042 4150 4	0·088 5699 $\int_9^8$
	2	$+4I-12-33+24$	0·167 7584	$\int_9^2$ 0 4392 8455 8212 2	0 103 7667 $\int_9^7$
	4	$-1I+22-33+44$	0·022 9217	$\int_9^4$ 0·2001 4707 8292 3	0·154 3037 $\int_9^5$

10	1	$+11+33$	0 429	7741	$\int_{10}^{1}$	0·9380	9428	7032	89	0 078	5463	$\int_{10}^{9}$
	3	$+31-13$	0 088	3712	$\int_{10}^{3}$	0 2825	3219	4188	278	0·105	7899	$\int_{10}^{7}$
11	1	$-21-42+53+34+15$	0 436	3033	$\int_{11}^{1}$	0 9431	6568	3002		0·070	5593	$\int_{11}^{10}$
	2	$+41-32-13-54+25$	0 184	1775	$\int_{11}^{2}$	0·4483	0768	3777		0 079	9529	$\int_{11}^{9}$
	3	$+51+12-43-24+35$	0 097	0280	$\int_{11}^{3}$	0 2859	7918	5905		0·091	9230	$\int_{11}^{8}$
	4	$-31+52+23-14+45$	0 049	3630	$\int_{11}^{4}$	0 2063	4948	9256		0 107	6230	$\int_{11}^{7}$
	5	$+11-22+33-44+55$	0 015	2762	$\int_{11}^{5}$	0·1595	4426	3923		0 128	9919	$\int_{11}^{6}$
12	1	$+11+33+55$	0 441	7121	$\int_{12}^{1}$	0 9474	6965	4203		0 064	0456	$\int_{12}^{12}$
	5	$+51-33+15$	0 026	2285	$\int_{12}^{5}$	0 1617	4591	5929		0 109	2889	$\int_{12}^{7}$
13	1	$-21-42-63+54+35+16$	0 446	2679	$\int_{13}^{1}$	0 9511	6803	2799	66	0 058	6322	$\int_{13}^{13}$
	2	$-41+52+13+34-65-26$	0 194	9966	$\int_{13}^{2}$	0 4550	0194	8695	2	0 065	0088	$\int_{13}^{11}$
	3	$-61+12+53-24-45+36$	0 109	4302	$\int_{13}^{3}$	0 2916	4461	3128	6	0 072	7842	$\int_{13}^{10}$
	4	$-51-32+23+64-15+46$	0 064	3735	$\int_{13}^{4}$	0 2111	9348	6757		0 082	4465	$\int_{13}^{9}$
	5	$+31-62-43+14-25+56$	0 034	4955	$\int_{13}^{5}$	0 1637	2371	6462		0 094	7328	$\int_{13}^{8}$
	6	$-11+22-33+44-55+66$	0 010	9094	$\int_{13}^{6}$	0 1326	2744	4748	7	0 110	8086	$\int_{13}^{7}$
14	1	$+11+33+55$	0·450	1581	$\int_{14}^{1}$	0 9543	8016	3579		0 054	0630	$\int_{14}^{13}$
	3	$+31-53-15$	0 114	0456	$\int_{14}^{3}$	0·2940	0000	1462		0 065	9088	$\int_{14}^{11}$
	5	$+51-13+35$	0 040	9773	$\int_{14}^{5}$	0 1655	0958	5205	2	0 083	5551	$\int_{14}^{9}$

d	a	${}_d\delta^a - \frac{2}{d}$				${}^4_dS^a_3$				${}^4_dS^{d-a}_3$			
15	1	$-21-42-63+74+55+36+17$	0 453	5218	$\int_{15}^{1}$	0 9571	9598	0986		0·050	1523	$\int_{15}^{14}$	
	2	$+41-72-33-14-55+66+27$	0 202	7009	$\int_{15}^{2}$	0 4601	6412	3884		0 054	7624	$\int_{15}^{13}$	
	4	$-71-12+63+24-55-36+47$	0 074	1629	$\int_{15}^{4}$	0 2150	7719	0077		0 066	7514	$\int_{15}^{11}$	
	7	$+11-22+33-44+35-66+77$	0 008	1812	$\int_{15}^{7}$	0 1134	7774	2752		0 097	1154	$\int_{15}^{8}$	
16	1	$+11+33+55+77$	0 456	4573	$\int_{16}^{1}$	0 9596	8453	8624		0·046	7699	$\int_{16}^{15}$	
	3	$+31-73-15-57$	0 121	2861	$\int_{16}^{3}$	0 2979	9606	9004		0 055	4239	$\int_{16}^{13}$	
	5	$+51-13+75+37$	0 050	5322	$\int_{16}^{5}$	0 1686	0602	0316		0·067	5416	$\int_{16}^{11}$	
	7	$+71-53+35-17$	0 014	5573	$\int_{16}^{7}$	0 1146	5540	2058		0·085	5409	$\int_{16}^{9}$	

Explanation. The 3rd and 4th columns give, respectively, the composition and numerical value of the difference term, ${}_d\delta^a$, which, added to or subtracted from $\frac{1}{2}{}^4_dS^a_D \left(= \frac{\pi}{2d} \operatorname{cosec} \frac{a\pi}{d} \right)$, gives, respectively, ${}^4_dS^a_3$ and ${}^4_dS^{d-a}_3$

The numbers in the 3rd column refer to that multiple of the angle π/d The first number denotes the cosine, which is multiplied by the second, which denotes the log of the secant, when in italics, and of the cotangent of half the angle, when in black type.

Examples $(d=5)-21+12$ is $-\cos\frac{2\pi}{5}\log_\epsilon \sec\frac{\pi}{5}+\cos\frac{\pi}{5}\log_\epsilon \sec\frac{2\pi}{5}$, $(d=8)+31-13$ is $+\cos\frac{3\pi}{8}\log_\epsilon \cot\frac{\pi}{16}-\cos\frac{\pi}{8}\log_\epsilon \cot\frac{3\pi}{16}$.

REFERENCES

Chrystal, G. 1886 *Text book of algebra*, part 1, p. 486. London: A. and C. Black.
Glaisher, J. W. L. 1873 *Quart. J. Math.* **12**, 232.
Hobson, E. W. 1928*a* *A treatise on plane trigonometry*, 7th ed. p. 371, cap XVII, example 25. Cambridge University Press.
Hobson, E. W. 1928*b* *A treatise on plane trigonometry*, 7th ed. p. 317, cap. XV, example 22. Cambridge University Press.
Soddy, F. 1936 *Nature, Lond.*, **137**, 1021 and **138**, 958.
Soddy, F. 1937 *Nature, Lond.*, **139**, 77.
Soddy, F. 1942 *Proc. Roy. Soc.* A, **179**, 377.

Relaxation methods applied to engineering problems. VIII. Plane-potential problems involving specified normal gradients

BY R V SOUTHWELL, F.R.S. AND GILLIAN VAISEY

(*Received* 11 *February* 1943)

Since every plane-harmonic function is associated with a conjugate, problems in which normal gradients are specified on the boundary can be transformed into problems in which boundary values are specified. There then remains, however, the problem of deducing a function ψ from its conjugate ϕ, and thus, when the conjugate has been determined only approximately, entails uncertainties which were exemplified in Part V. To minimize the errors of approximate computation ψ and ϕ should be determined *severally and independently*, consequently a method of direct attack is still needed on problems in which normal gradients are specified. Recent applications have, moreover, presented cases in which the boundary conditions are 'mixed', i.e. values are specified at some parts of the boundary, gradients at others.

Here, two methods are propounded for the satisfaction of mixed boundary conditions, the first applicable also to cases in which normal gradients alone are specified. Test examples indicate that the wanted extension of method is now available.

INTRODUCTION

1. There are two standard cases of the plane-potential problem. In the first, the wanted function ψ is governed by Laplace's equation

$$\nabla^2\psi \equiv \frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2} = 0 \qquad (1)$$

and takes specified values at points on the boundary. This case was treated in Part III (Christopherson & Southwell 1938),* along with variants in which account

* The short title 'Part III' will be used throughout this paper.

was taken of refraction and plasticity, or in which (1) was replaced by Poisson's equation

$$\nabla^2 w + Z = 0, \tag{2}$$

Z being a specified function of x and y. The second case is different in that the normal gradient of the wanted function, instead of the function itself, has to take specified values at the boundary. In both cases the solution is *unique*, except that in the second any constant quantity may be added to w.

A particular integral of (2) can always be formulated, either exactly by a known formula in the theory of attractions, or approximately, after expression of $\nabla^2 w$ in finite differences, consequently any problem governed by an equation of type (2) can be reduced to a plane-potential problem governed by (1). Moreover with every solution of (1) there is associated a conjugate plane-harmonic function ϕ such that $(\phi + i\psi)$ is a function of the complex variable $(x + iy)$,—i.e. such that

$$\left.\begin{aligned} \frac{\partial\phi}{\partial x} &= \frac{\partial\psi}{\partial y}, & -\frac{\partial\phi}{\partial y} &= \frac{\partial\psi}{\partial x}, \\ \text{and therefore} \quad \frac{\partial\phi}{\partial \nu} &= \frac{\partial\psi}{\partial s}, & -\frac{\partial\phi}{\partial s} &= \frac{\partial\psi}{\partial \nu}, \end{aligned}\right\} \tag{3}$$

when s, measured along the boundary, and ν, the outward normal, are related as in figure 1. This means that a problem of the second class, in which $\partial\psi/\partial\nu$ has specified values at the boundary, can be formulated and solved as a problem of the first class, in which $\partial\phi/\partial s$ (and consequently ϕ, except for an arbitrary constant) has specified values. The subsequent problem of deriving ψ from ϕ was discussed in Part V (Gandy & Southwell 1940).*

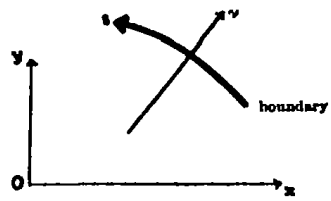


FIGURE 1

2. Thus the two standard classes can be identified, and from an 'orthodox' standpoint they do not call for separate discussion. But a third class of problem can be envisaged which is *not* similarly reducible either to the first or to the second, viz. that class in which the wanted function is governed by 'mixed' boundary conditions, its value being specified at some, its normal gradient at other points on the boundary. We have had no occasion, so far, to discuss it in this series, but a recent application of the relaxation method requires that it now be considered. Plainly its solutions will be unique, like those in which only boundary values are specified.

Moreover the second class too calls for examination from a 'relaxation' standpoint, because we must expect as unavoidable errors due to incomplete 'liquidation of residuals'. Strictly speaking, unless a function satisfies equation (1) exactly it has no conjugate, and on this account an element of arbitrariness is apparent in the methods of Part V, whereby having determined ψ *approximately* we sought to

* The short title 'Part V' will be used throughout this paper.

determine its conjugate function ϕ. Here it is shown (§ 5) that if we impose as the condition for 'optimal conjugacy' of ψ and ϕ that the integral

$$Q = \iint \left\{ \left(\frac{\partial \phi}{\partial x} - \frac{\partial \psi}{\partial y} \right)^2 + \left(\frac{\partial \phi}{\partial y} + \frac{\partial \psi}{\partial x} \right)^2 \right\} dx dy \tag{4}$$

—evaluated for the whole region contained within the specified boundary—must take its minimal value, then when ψ is given ϕ will be subject to the conditions

$$\left.\begin{aligned} \nabla^2 \phi = 0, &\quad \text{everywhere in the region,} \\ \frac{\partial \phi}{\partial \nu} = \frac{\partial \psi}{\partial s}, &\quad \text{everywhere on the boundary.} \end{aligned}\right\} \tag{5}$$

This means that whether or no ψ has been computed, strictly speaking we ought to compute ϕ independently, by a technique devised specially for problems of the second class.

3. We are thus impelled to search for such technique, and it should if possible be also applicable to 'mixed' boundary conditions. This is the purpose of the present paper, which propounds for mixed boundary conditions two methods of attack—the first applicable also to cases in which normal gradients alone are specified (class 2).

The basis of the methods of Part III was that mechanical interpretation of (1) and (2) which is afforded by 'Prandtl's membrane analogue'. Seen from this standpoint, specified values of $\partial w/\partial \nu$ are specified values of $T(\partial w/\partial \nu)$, T denoting the tension in the membrane, i.e they are line-intensities of transverse loading applied to the membrane at its edge. 'Method 1' (§§ 11–13) adopts this physical picture and proceeds in the manner of Part III, replacing the continuous membrane by a net of finite mesh, and accepting the errors which result from such replacement (on the ground that these become insensible as advance is made to successively finer nets). It is equally applicable to problems in class 2 and in class 3 (§ 2). 'Method 2' (due in principle to Dr J. R. Green) is an alternative treatment of mixed boundary conditions (class 3) which seems likely to have—in special circumstances—some advantages over Method 1. It was tested (by Dr L. Fox) before 'Method 1' had been fully developed.

The worked examples were chosen solely as test cases, not as having any intrinsic interest. Their solutions are known, and without entailing inessential complications each leaves room for all errors really characteristic of the procedure tested.

4. Being concerned with fundamental aspects, this paper affords an opportunity to record some extensions of the theory first outlined in Part III. In Section I, § 5 deals with the problem noticed in § 2, namely, given an approximation to some plane-harmonic function ψ, to obtain the best approximation to its conjugate function ϕ; §§ 6–9 give a physical explanation of the fact (noticed in Part III, but only as a mathematical deduction) that closer approximation is attainable with a use of triangular nets ($N = 6$) than with either square or hexagonal nets ($N = 4$ or 3);

§ 9 completes our physical pictures of 'relaxation nets' regarded as actual nets of tensioned strings transversely loaded at their nodal points,—showing what string tension must be postulated in order that a chosen net may be the finite-difference approximation to the tensioned membrane of 'Prandtl's analogue'.

Section I. Some general theorems

Approximate computation of conjugate plane-harmonic functions

5. When (owing to unavoidable errors of computation) ψ is not strictly plane-harmonic, no single-valued function ϕ will satisfy *both* of

$$\frac{\partial\phi}{\partial x}=\frac{\partial\psi}{\partial y}, \qquad -\frac{\partial\phi}{\partial y}=\frac{\partial\psi}{\partial x}, \tag{3 bis}$$

and we have the problem of choosing ϕ to minimize the residual error. We may take as the measure of this error the integral

$$Q=\iint\left\{\left(\frac{\partial\phi}{\partial x}-\frac{\partial\psi}{\partial y}\right)^2+\left(\frac{\partial\phi}{\partial y}+\frac{\partial\psi}{\partial x}\right)^2\right\}dxdy, \tag{4 bis}$$

evaluated for the whole 'field' contained within the specified boundary. i.e., given ψ we may say that the optimal approximation to its conjugate function ϕ is that which makes Q a minimum.

Let
$$\phi=\phi_0+\phi', \tag{6}$$

where ϕ_0 satisfies the conditions

$$\left.\begin{aligned}&\nabla^2\phi=0, \quad \text{everywhere in the 'field',}\\ &\frac{\partial\phi}{\partial\nu}=\frac{\partial\psi}{\partial s}, \quad \text{everywhere on the boundary.}\end{aligned}\right\} \tag{5 bis}$$

Then on substituting in (4) from (6) we have

$$\begin{aligned}Q=&\iint\left\{\left(\frac{\partial\phi_0}{\partial x}-\frac{\partial\psi}{\partial y}\right)^2+\left(\frac{\partial\phi_0}{\partial y}+\frac{\partial\psi}{\partial x}\right)^2\right\}dxdy+\iint\left\{\left(\frac{\partial\phi'}{\partial x}\right)^2+\left(\frac{\partial\phi'}{\partial y}\right)^2\right\}dxdy\\ &+2\iint\left\{\frac{\partial\phi'}{\partial x}\left(\frac{\partial\phi_0}{\partial x}-\frac{\partial\psi}{\partial y}\right)+\frac{\partial\phi'}{\partial y}\left(\frac{\partial\phi_0}{\partial y}+\frac{\partial\psi}{\partial x}\right)\right\}dxdy,\end{aligned} \tag{7}$$

and the last of the integrals in this expression

$$=2\oint\phi'\left(\frac{\partial\phi_0}{\partial\nu}-\frac{\partial\psi}{\partial s}\right)ds-2\iint\phi'\nabla^2\phi_0\,dxdy, \quad \text{by Green's transformation,}$$

$=0$ since ϕ_0 satisfies the relations (5).

Again, since ϕ_0 and ψ are invariant, so also is the first integral in (7). Consequently Q has its minimum value when

$$\iint\left\{\left(\frac{\partial\phi'}{\partial x}\right)^2+\left(\frac{\partial\phi'}{\partial y}\right)^2\right\}dxdy=0,$$

i.e. when $\phi'=$ const., so that ϕ (like ϕ_0) satisfies (5).

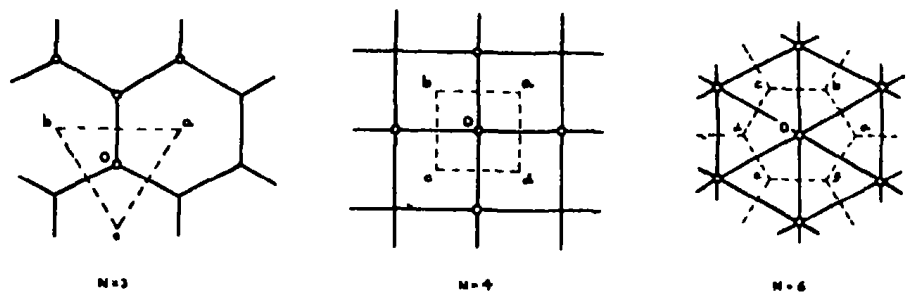
According to this conclusion *the optimal approximation to ϕ is governed by* (5) *and as such is independent of any errors in the computation of ψ*. Its determination is a plane-harmonic problem of the second class (§ 1), calling for special technique.

Physical aspects of the relaxation net

6. In Part III* it was shown (§§ 6-8) that when $N=3$ or 4

$$\left.\begin{aligned}&\frac{1}{N}\Sigma_{a,N}(w)-w=\frac{a^2}{4}\nabla^2w=-Z\frac{a^2}{4}, \quad \text{according to (2),} \quad \text{(i)}\\ &\text{when } N=6\\ &\frac{1}{N}\Sigma_{a,N}(w)-w=\left(\frac{a^2}{4}\nabla^2w+\frac{a^4}{64}\nabla^4w\right)=-\left(Z\frac{a^2}{4}+\nabla^2Z\frac{a^4}{64}\right),\\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{according to (2),} \quad \text{(ii)}\end{aligned}\right\} \tag{8}$$

with neglect in every instance of terms of order a^N. (Here w, ∇^2w, ∇^4w are values taken at some nodal point O of a regular net (figure 2), round which point N other points are arranged symmetrically at a distance a. $\Sigma_{a,N}(w)$ stands for the sum of the values of w at these N surrounding nodes)



FIGURE

The difference of the left- and right-hand sides of (i) or (ii) was taken as a measure of the 'residual force' at O, and rules were devised whereby all residual forces can be brought to zero by a 'systematic relaxation of constraints'. For computation it has been found convenient to multiply this measure of residual force by N, i.e. to replace (i) and (ii) of (8) by

$$\mathbf{F}=F+\mathsf{F}=0, \tag{9}$$

* In Part III, Z was written for ∇^2w, i.e. for $-Z$ as employed in (2) and throughout this paper.

where F, the 'force' exerted on account of 'displacement' w,

$$= \Sigma_{a,N}(w) - Nw,$$

$$\mathsf{F}, \text{ the 'external force'}, = -\frac{Na^2}{4}\nabla^2 w = NZ\frac{a^2}{4}, \quad \text{when } N = 3 \text{ or } 4, \tag{10}$$

$$= N\left(Z\frac{a^2}{4} + \nabla^2 Z\frac{a^4}{64}\right), \quad \text{when } N = 6.$$

We shall adopt this modified convention here, taking (9) and (10) as the basis of our approximate treatment of the governing equation (2), § 1.

7. A mechanical interpretation of (2) is given by 'Prandtl's membrane analogue': it governs the small transverse displacement w of a membrane strained to uniform tension T and loaded with transverse pressure of surface-intensity TZ (Z being a function of x and y). A corresponding interpretation of (9) and (10) was propounded in Part III, § 9. they govern the small transverse displacement w, due to transverse forces applied at its nodal points, of a net, initially flat, in which every string is strained to the same tension T.

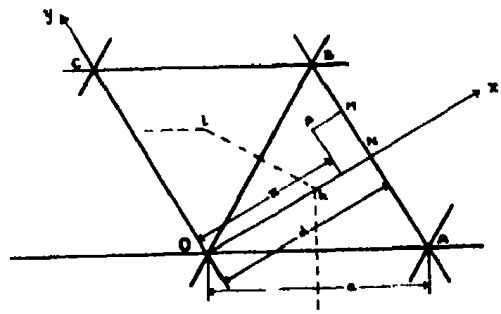


FIGURE 3

To complete this physical picture we must account for the magnitudes of the loading (Z) terms in (10), and in particular for the additional term which appears when $N = 6$. Clearly, if a distributed loading is to be brought to bear upon a net of finite mesh, and in such a way that every string connecting two adjacent nodes remains straight when the net is deflected, some mechanism must be envisaged for effecting the distribution, e.g a light rigid plate having the shape of the mesh and suspended from its nodal corners by inextensible ties.* Suppose, in the first place, that the mesh is triangular ($N = 6$) and that a concentrated force Z acts vertically through the point P of the rigid plate OAB (figure 3). Then the part of Z which on statical principles should be sustained by O is $\mathrm{Z}(PM/ON)$, i e. it is $\mathrm{Z}(1 - x/d)$, where $d = \frac{1}{2}a\sqrt{3}$. Dealing in this manner with the pressure TZ over the whole of OAB, we find that the total contribution of the pressures on this triangle to the force at O is

$$T\iint Z\left(1 - \frac{x}{d}\right)dxdy, \quad \text{for the whole triangle } OAB, \tag{i}$$

* Or furnished with feet which bear upon the nodal points.

and if in the manner of Part III, §§ 6–7, we now write

$$\begin{aligned} Z = Z_0 &+ \tfrac{1}{4}(\nabla^2 Z)_0 (x^2+y^2) + \ldots \\ &+ Ax + By + F(x^2-y^2) + 2Gxy \\ &+ \ldots \text{ (terms of higher order in } x, y), \end{aligned} \tag{ii}$$

then, to the order of the terms retained in (ii), we have in (i)

$$\iint\left(1-\frac{x}{d}\right) Z\,dx\,dy = (\text{area } OAB) \times [\tfrac{1}{3}Z_0 + \tfrac{1}{6}Ad + \tfrac{1}{10}Fd^2(1-\tfrac{1}{5}) + \tfrac{1}{40}(\nabla^2 Z)_0 d^2(1+\tfrac{1}{5})]. \tag{iii}$$

For axes Ox', Oy' inclined at θ to Ox, Oy, so that

$$x = x'\cos\theta - y'\sin\theta, \quad y = x'\sin\theta + y'\cos\theta,$$

the expression (ii) would become

$$\begin{aligned} Z = Z_0 &+ \tfrac{1}{4}(\nabla^2 Z)_0 (x'^2+y'^2) + \ldots \\ &+ (A\cos\theta + B\sin\theta)x' + (B\cos\theta - A\sin\theta)y' \\ &+ (F\cos 2\theta + G\sin 2\theta)(x'^2-y'^2) \\ &\qquad + 2(G\cos 2\theta - F\sin 2\theta)x'y' \\ &+ \quad \text{(terms of higher order in } x', y'), \end{aligned}$$

so for the triangular mesh OBC adjacent to OAB (i e a triangle like OAB but rotated through $\frac{1}{3}\pi$) the expression corresponding with (iii) is

$$\begin{aligned} (\text{area } OAB) \times [\tfrac{1}{3}Z_0 &+ \tfrac{1}{6}(A\cos\tfrac{1}{3}\pi + B\sin\tfrac{1}{3}\pi)d \\ &+ \tfrac{5}{60}d^2(F\cos\tfrac{2}{3}\pi + G\sin\tfrac{2}{3}\pi) + \tfrac{1}{360}(\nabla^2 Z)_0 d^2] \end{aligned}$$

The other four triangles surrounding O can be treated similarly Summing the contributions of all six triangles, we find that the terms involving A, B, F, G sum to zero, leaving

$$T \times (\text{area } OAB) \times 2[Z_0 + \tfrac{1}{12}(\nabla^2 Z)_0 d^2] \tag{iv}$$

as the total force which, according to (i), comes on the nodal point O. Since $4d^2 = 3a^2$ and since the area $OAB = \frac{\sqrt{3}}{4}a^2$, the expression (iv) may be written as

$$\mathsf{F} = T\frac{\sqrt{3}a^2}{2}\left[Z_0 + \frac{a^2}{16}(\nabla^2 Z)_0\right], \tag{v}$$

and this (for $N = 6$) agrees with F as given by the last of (10), provided that

$$T = \sqrt{3}. \tag{11}$$

8. T denoting the tension in any one string, it is easy to show that

$$F = \frac{\mathrm{T}}{a}\{\Sigma_{a,N}(w) - Nw_0\}$$

is the force which acts at O in consequence of small displacements (w) of the surrounding nodes. This (for $N = 6$) agrees with F as given by the first of (10), provided that

$$\mathbf{T} = a = aT/\sqrt{3}, \quad \text{when (11) is satisfied,}$$

$$= T \times a \tan \pi/N, \tag{12}$$

i.e. *provided that we concentrate in the string OB (figure 3) the whole force which in the membrane was exerted by the tension T acting across the line $kl = (OB)\tan \pi/N$.* (On our assumption of a transverse pressure TZ, it was evident that the magnitude of T would be immaterial.)

9. Our physical picture of triangular nets is now complete. Equations (9) and (10) express the condition for transverse equilibrium of nodal points when

(i) every string tension $\mathbf{T}$ has a value equivalent to the resultant pull of the membrane tension T ($= \sqrt{3}$) across a side ($a \tan \pi/N$) of the hexagon surrounding that point,

(ii) the transverse pressure (TZ) is concentrated at nodal points in accordance with the principles of Statics.

(They will hold when F and F are multiplied by any factor; but no factor is involved when $T = \sqrt{3}$ and $\mathbf{T}$ accordingly $= a$.)

But when the mesh is square or hexagonal ($N = 4$ or 3), statical principles no longer suffice to determine the partition of the transverse pressure between adjacent nodal points, and the best that we can do is to concentrate at O (figure 2) the whole of the pressure acting on the surrounding polygon $abc \ldots a$. On that understanding the total pressure-force at O is

$$TZ_0 \times (\text{area } abc \ldots a) = TZ_0 \times N(\tfrac{1}{2}a \times \tfrac{1}{2}a \tan \pi/N)$$

$$= \tfrac{1}{4} N Z_0 T a^2 \tan \pi/N \tag{i}$$

If moreover (as in § 8) we concentrate in each string the total force exerted by the membrane tension T acting across a side of the polygon $abc \ldots a$, then we have

$$F = \frac{\mathbf{T}}{a}\{\Sigma_{a,N}(w) - Nw_0\}, \quad \text{where } \mathbf{T} = Ta \tan \pi/N,$$

as in that section. Combining this result with (i), we deduce that the residual force at O is

$$\mathbf{F} = T \tan \pi/N\{\tfrac{1}{4} N Z_0 a^2 + \Sigma_{a,N}(w) - Nw_0\}, \tag{13}$$

and this, if we now make

$$T = \cot \pi/N, \tag{14}$$

is equivalent to (9) and (10). For $N = 6$ the assumption (14) is equivalent to (11) of § 7.

10. To summarize these conclusions· Equations (9) and (10) may be employed as the finite-difference equivalent of (2); and just as (2), in 'Prandtl's membrane

analogue', is interpreted as governing the (small) transverse displacement w of a membrane strained to uniform tension T and loaded by pressure of surface-intensity TZ, so (9) and (10), in our 'net analogue', can be interpreted as governing the nodal displacements w of a net in which every string has the same tension $\mathbf{T}$ and nodes sustain transverse forces F as given by the second of (10). For correspondence of displacements in the membrane and in the net, $\mathbf{T}$ must be related with T by (12) and F must be, *for $N = 6$ the statical equivalent, for $N = 3$ or 4 the resultant*, of TZ acting over a certain area. We must, moreover, have

$$T = \cot \pi/N, \qquad (14)\,bis$$

and hence, according to (12),

$$\mathbf{T} = a, \qquad (15)$$

in order that the transverse force exerted by any one string in consequence of a unit difference between the values of its terminal displacements may be $\mathbf{T}/a = 1$. (In Part III, written before multiplication by N (§ 6) had been found convenient, the corresponding force was $1/N$.)

When, in place of (2), Laplace's equation (1) is to be satisfied, Z does not enter into our equations, and then in problems of the first class (§ 1) —i.e. when boundary values are specified—this close study of numerical factors is not necessary; but it is needed in problems of the second or third class —i e when normal gradients are specified. In the membrane analogue (cf. § 3) specified values of $\partial w/\partial \nu$ are specified values of $T(\partial w/\partial \nu)$, the line-intensity of a transverse edge-loading. using the 'net approximation', we shall (in Method 1 cf § 11) concentrate this distributed loading on the strings which cross the boundary, and then in order to proceed we must know the consequent slope of every such string. Using (9) and (10) we assume that *a unit transverse force on any string entails a unit difference between the displacements of its two ends*· then according to (12) $T = \cot \pi/N$, and so T must be given this value in integrals of the type $\int T \dfrac{\partial w}{\partial \nu} ds$, which measure the transverse forces.

Section 11. The treatment by relaxation methods of specified normal gradients

Method 1

11 As shown in § 10, the total transverse edge-loading to be taken as acting on an arc of boundary is the value of

$$\int T \frac{\partial w}{\partial \nu} ds = \cot (\pi/N) \int \frac{\partial w}{\partial \nu} ds \qquad (16)$$

for that arc, when (9) and (10) apply. Given $\partial w/\partial \nu$ at every point on the boundary, by graphical and/or numerical methods we can compute the integral for any selected arc.

In Method 1 we make this computation for every element of boundary which is intercepted by 'strings' of the chosen net, and we concentrate each force (16), in

accordance with the rules of Statics, on the strings which terminate the relevant arc. Thereby we derive a problem suited to computation by relaxation methods, in which strings crossing the boundary have not there (as in Part III) to undergo specified displacements, but to sustain specified transverse forces.

The derived forces sum to zero, because for equilibrium of the original (specified) membrane it is essential that*

$$T\oint \frac{\partial w}{\partial \nu} ds \left(= T\iint \nabla^2 w \, dx\, dy \right) = 0. \tag{17}$$

Our problem in relaxation will be to make them 'meet' and cancel (whereas in problems of the first class, treated in Part III, we had only to transfer them to the boundary).

12. Concentrating the edge-loading in this way, without affecting the equilibrium of internal nodes we may imagine each loaded string to be prolonged beyond the boundary and its load Z to be transferred to its outer end (D, figure 4). As in Parts III and VII, we may conveniently assume every string to have the standard length, so that its outer end becomes a 'fictitious node' outside the boundary. The question then arises, whether the fictitious nodes should or should not be assumed to have strings connecting them: the answer would seem to be, that in some circumstances strings should be assumed to connect fictitious nodes, but that then their tensions should have half the standard value, in other circumstances the fictitious nodes should *not* be so connected.

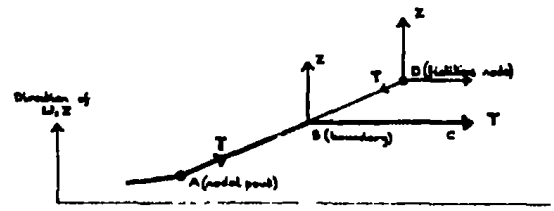


FIGURE 4

The argument may be stated as follows: The fictitious nodes, if we join them, will define a polygonal boundary lying either on or outside the specified boundary, so that a membrane of specified plan-form (the 'Prandtl analogue' of our plane-harmonic problem) will be replaced in our treatment by a slightly larger not representing (approximately) a slightly larger polygonal membrane, indicated by shading in Fig. 5a. But now the tension in a fictitious string comes (§ 8) from the membrane tension in an element of which half lies inside, half outside the boundary. Here the outer half has no real existence (figure 5a), so we must make the tension $\frac{1}{2}$T.

On the other hand, plainly the polygonal (substituted) membrane should—for

* We observe that this condition necessitates a treatment which (like Method 1) deals with *integrated* effects of the transverse loading.

accurate results—be no larger than is necessary, and sometimes this consideration will suggest the replacement indicated in figure 5*b*. There the shaded area indicates a membrane larger than what is specified, but lying within the polygon (indicated by broken lines) which would be obtained by joining fictitious nodes in the manner of figure 5*a*. Accepting this shaded area as a substitute for the specified membrane, and now replacing it by a net, we shall account for all of the membrane tension by concentrating a tension T in each of the strings (AB, ..., etc.) which join fictitious nodes to nodes inside the boundary. *There is no longer any membrane element calling for a string to connect fictitious nodes, so no tension is to be associated with any of the broken lines.*

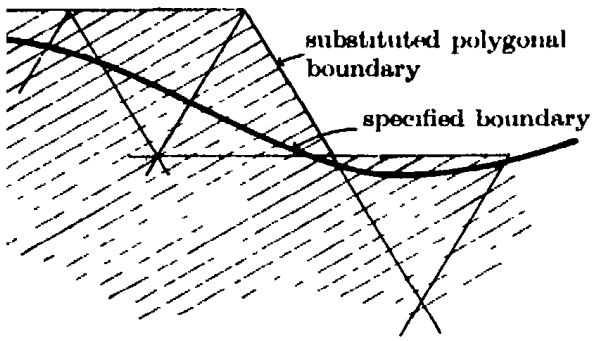


FIGURE 5*a*

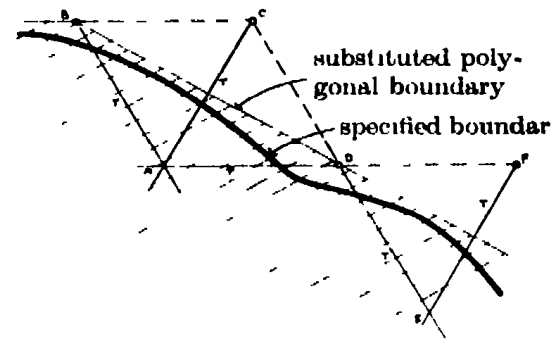


FIGURE 5*b*

13. Figs. 5 relate to triangular nets; but nets of square mesh can be treated similarly, and in either instance we are left with a very simple problem of liquidation, entailing no departure from the standard 'relaxation pattern' excepting where (at the polygonal boundary) the strings are less than N in number and/or exert 'half tensions' as explained in § 12. Complete liquidation will leave no residual force at any node, whether inside or on the modified boundary. It is possible (theoretically) because the initial forces must sum to zero, as shown in § 11.

The technique of liquidation has been described already, in Parts III and V.

Example 1

14. To test the method we have applied it to a problem specially chosen (i) as having a known solution, simple to compute, and (ii) as entailing rapid variations of $\partial w/\partial v$ in some part of the boundary.

The function
$$w = \theta = \tan^{-1}\frac{y}{x} \tag{18}$$

is plane-harmonic and can be evaluated for any point (x, y) from a table of inverse tangents; and on a circular boundary touching the axes of co-ordinates, viz.

$$x^2+y^2-2R(x+y)+R^2 = 0, \tag{19}$$

its normal gradient

$$\frac{\partial w}{\partial \bar{\nu}} = \left(\frac{\partial w}{\partial \bar{r}}\right)_{r=R}, \quad \text{where } r^2 = (x-R)^2 + (y-r)^2,$$

$$= \left[\frac{\partial}{\partial \bar{r}} \tan^{-1}\left(\frac{R + r\sin\phi}{R + r\cos\phi}\right)\right]_{r=R}$$

$$= \frac{1}{R}\left(\frac{\sin\phi - \cos\phi}{3 + 2\sin\phi + 2\cos\phi}\right), \quad \text{where } \tan\phi = \frac{y-R}{x-R}. \tag{20}$$

Figure 6 explains the significance of θ, R, r, ϕ.

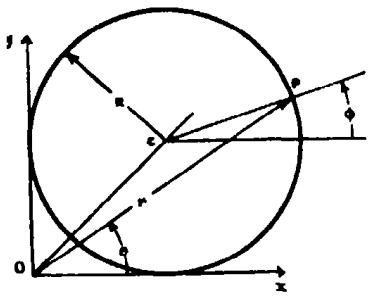


FIGURE 6

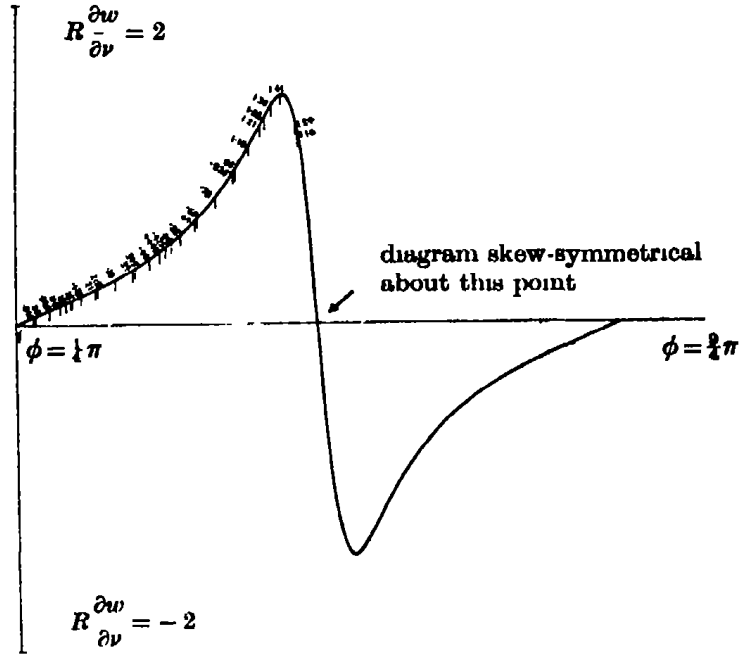

$R\frac{\partial w}{\partial \nu} = 2$
diagram skew-symmetrical about this point
$\phi = \frac{1}{4}\pi$
$\phi = \frac{9}{4}\pi$
$R\frac{\partial w}{\partial \nu} = -2$

FIGURE 7

Figure 7 exhibits the variation with ϕ of $R(\partial w/\partial \nu)$ as given by (20). The variation is rapid in the neighbourhood of $\phi = \frac{5}{4}\pi$; so by taking as our first example the determination of a plane-harmonic function w which on the circular boundary (19) has normal gradients given by (20) we shall be imposing a severe test on our computational method, notwithstanding that the wanted function is in fact simple (its contours being straight lines through the origin).

The test moreover is fair, because the data (i.e. the forces to be concentrated on the several strings) can be calculated and so are not subject to inaccuracies resulting from faulty estimation of normals According to (20)

$$\int \frac{\partial w}{\partial \bar{\nu}}\, ds = R\int \frac{\partial w}{\partial \bar{\nu}}\, d\phi = \int \frac{\sin\phi - \cos\phi}{3 + 2\sin\phi + 2\cos\phi}\, d\phi$$

$$= -\tfrac{1}{2}\log(3 + 2\sin\phi + 2\cos\phi), \tag{21}$$

the lower limit of integration being arbitrary; so $\oint \frac{\partial w}{\partial \nu} ds = 0$, as required in § 11. Figure 8 exhibits $\int \frac{\partial w}{\partial \nu} ds$ as a function of ϕ, with the lower limit taken at $\phi = \frac{1}{4}\pi$.

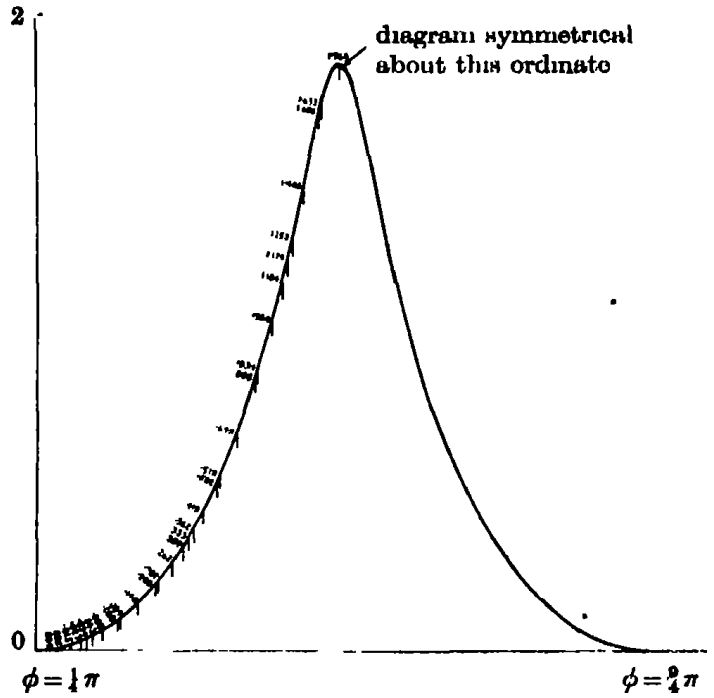


FIGURE 8

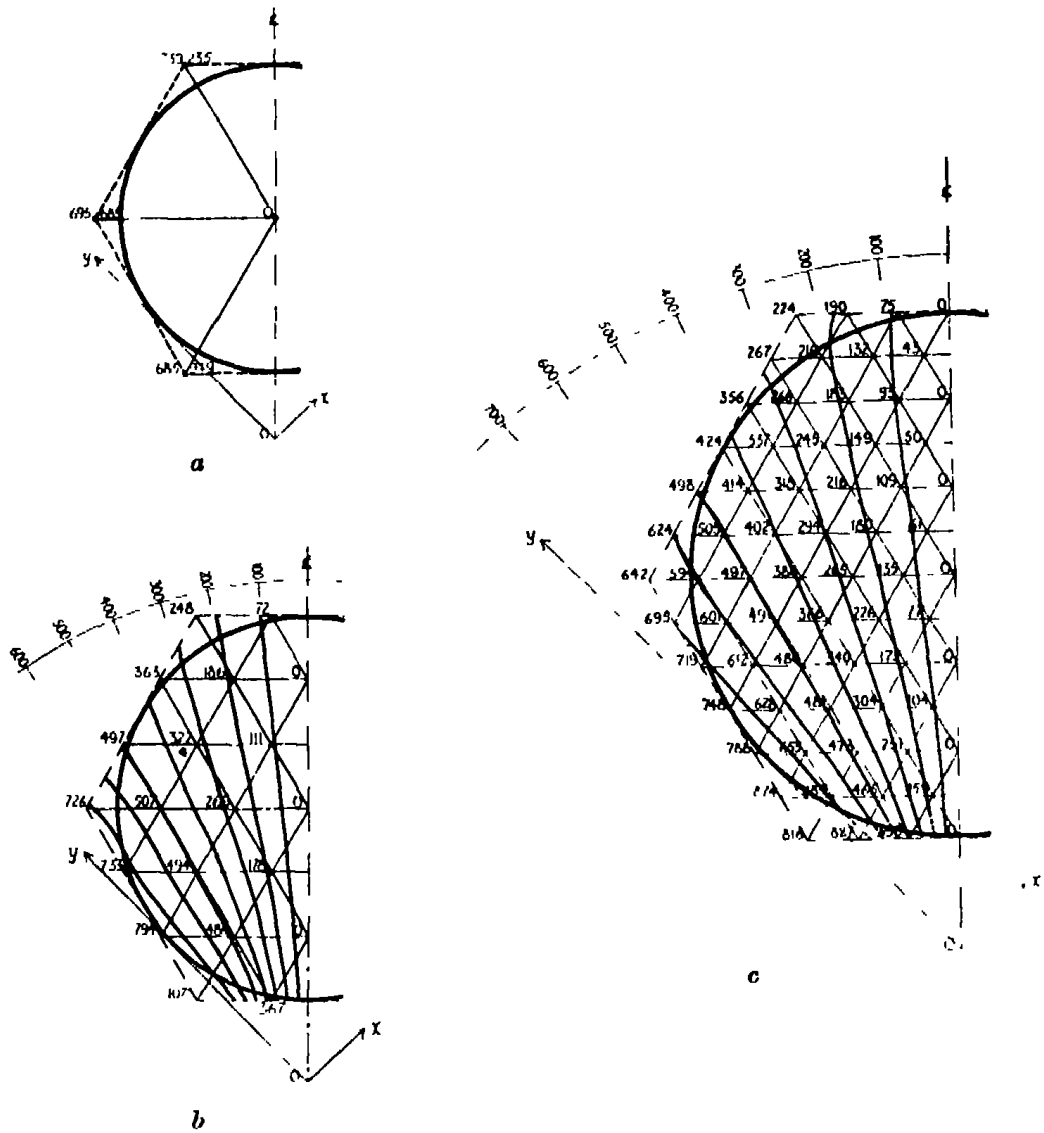
15. In figure 8, values of $\int \frac{\partial w}{\partial \nu} ds$ are recorded for estimated cutting points which are shown; the difference of any two adjacent values, multiplied by $\cot \frac{1}{6}\pi$ $(= \sqrt{3})$ in accordance with (16), gives the total load on the corresponding element of boundary, and the proportion to be concentrated at each end-point was estimated in accordance with statical principles (§ 11) Thereby we ensured that errors of estimation would leave forces still summing to zero (§ 11) and—on a fine net—departing so little from exact statical equipollence with the specified edge-loading as to entail no sensible error in the result.

Figure 9 shows, for three successive nets, the computed deflexions and (figures *b* and *c*)* contours for comparison with straight lines derived from the known solution. (These last, for clarity, are not drawn in the diagrams, but their directions are shown on circular arcs. All, of course, pass through the origin of co-ordinates.) To save space we have plotted, not w as given by (18), but w as given by

$$w = \theta - \tfrac{1}{4}\pi, \tag{22}$$

* In figure 9*a* exact values (underlined) are included for comparison. The net is too coarse to permit construction of contours.

which makes the wanted function skew-symmetrical with respect to OC in figure 6, therefore makes that line a contour ($w = 0$).* In each of the three diagrams, broken lines indicate strings (through fictitious nodes) which we assumed to exert 'half-tensions', chain lines with open circles (o) indicate strings which were given no tension (§ 12), i e were suppressed.



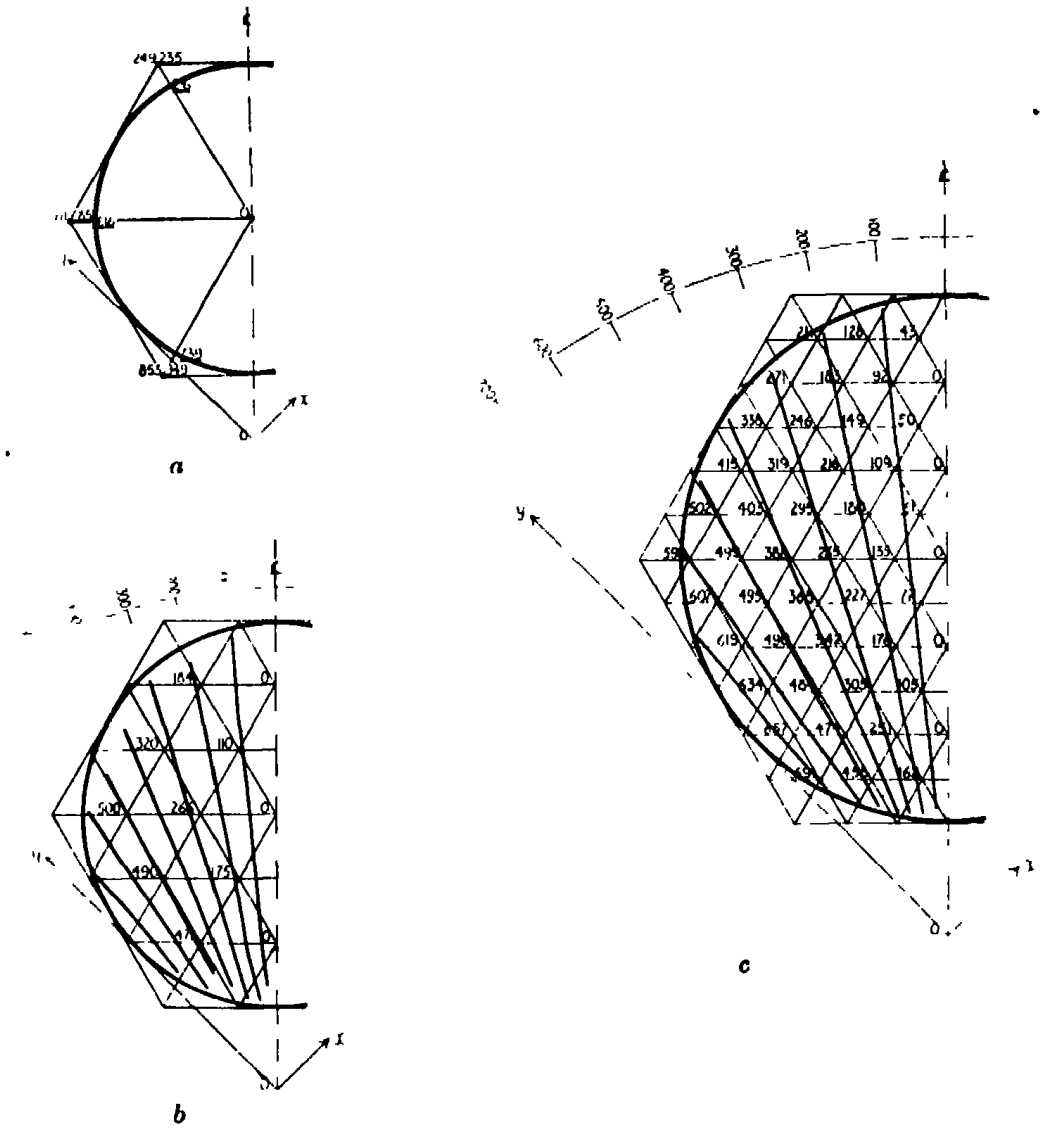
FIGURES 9. (Normal gradients specified)

16. The correspondence of our computed contours with the (correct) straight lines through the origin gives an indication of the over-all accuracy which can be achieved by Method 1. For comparison, figures 10 exhibit corresponding solutions, obtained by the methods of Part III, to the problem of determining w when boundary

* It was remarked in § 1 that any constant may be added to w in a problem of class 2.

values are specified (case 1 of § 1). There is not much difference in the accuracy attainable (with a net of given mesh-size) in the two cases.

The errors, though appreciable, are not of an order likely to matter seriously in physical applications, and are less than those obtained (e g.) by a use of 'electrical tanks' (cf. Bradfield, Hooker & Southwell 1937). They arise in part from our use of



FIGURES 10. (Boundary values specified)

finite-difference equations, because the wanted function (18), though plane-harmonic, does not exactly satisfy the finite-difference relation

$$\Sigma_{a,6}(w) - 6w = 0,$$

which is our analogue of Laplace's equation (1).

Example 2

17. Method 1 is also applicable on nets of square mesh, and to 'mixed' boundary conditions. A convenient example is provided by Saint-Venant's 'torsion-function' ϕ for an elliptical shaft. The plane harmonic function

$$\phi = -\frac{a^2-b^2}{a^2+b^2}xy \tag{23}$$

satisfies the condition $$\frac{\partial\phi}{\partial\nu} = y\cos(x,\nu) - x\cos(y,\nu) \tag{24}$$

at every point of a boundary represented by

$$\frac{x^2}{a^2}+\frac{y^2}{b^2} = 1, \tag{25}$$

and vanishes on the axis of co-ordinates; so its determination for a single quadrant ($x>0$, $y>0$, say) is a problem involving mixed boundary conditions. We here consider the case in which $a/b = \frac{5}{3}$, so that the required solution (23) is

$$\phi = -\tfrac{8}{17}xy \tag{26}$$

18. According to (16) and (24), the total load on an element of boundary is

$$\begin{aligned}\cot(\pi/N)\int\frac{\partial\phi}{\partial\nu}ds &= \cot(\pi/N)\int(y\,dy + x\,dx)\\ &= \tfrac{1}{2}(x^2+y^2)\cot(\pi/N),\end{aligned} \tag{27}$$

the lower limit of integration being arbitrary; on a single quadrant (§ 17) the total load is

$$\pm\tfrac{1}{2}(a^2-b^2)\cot(\pi/N).$$

From (27), having calculated the co-ordinates of points where strings cut the boundary, we can compute the load on each element of boundary, then concentrate this in accordance with statical principles. As in Example 1 (§ 14), slight errors of estimation entail—on a fine net—no sensible error in the result.

Figures 11 present solutions, on two sizes of net, of this example of mixed boundary conditions. In this instance (since ϕ according to (23) has linear variation along every string) the methods of Part III yield an *exact* result when boundary values are specified. The correct contours (rectangular hyperbolas) are indicated in fine lines for comparison.

Review of Method 1

19. Using known formulae of interpolation, it is possible to improve on the accuracy of our treatment of 'irregular stars', though at some cost in complexity; and a like result may be expected to follow from corresponding (i.e. mathematical) study of Method 1. But our aim in this series is to pursue as far as possible (for plane-harmonic problems) the notion of relaxation 'nets' regarded from a physical stand-

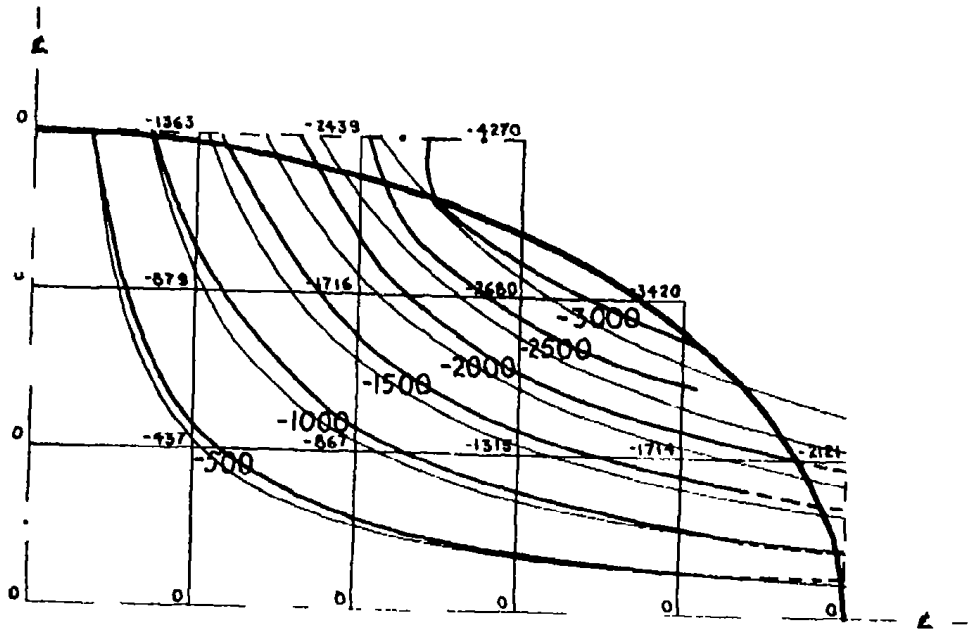


FIGURE 11*a*

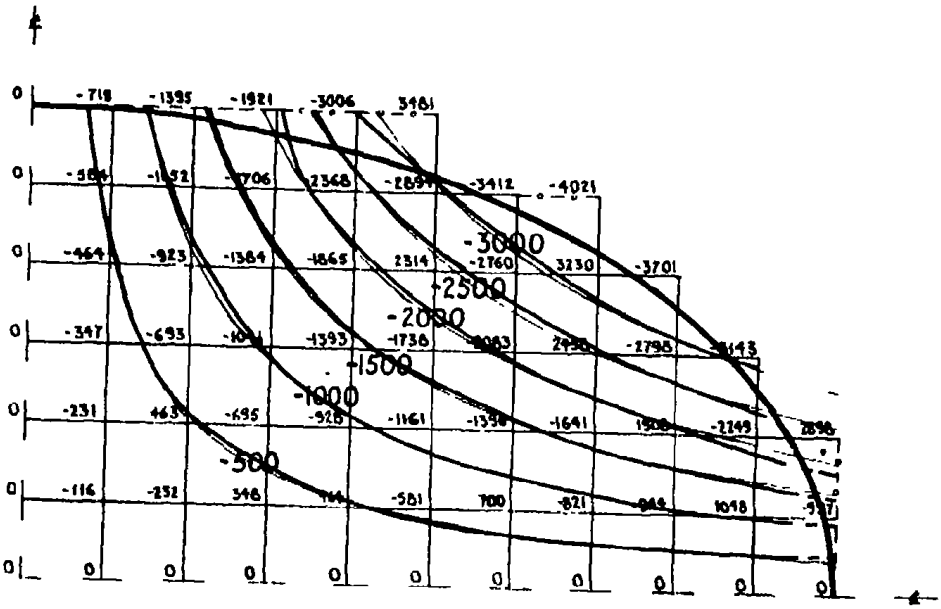


FIGURE 11*b*

point, i.e. as tensioned nets replacing the continuous membrane which is Prandtl's analogue (§ 3) of the governing (differential) equation. Accordingly (deeming it more important, at present, to extend the range than the accuracy of relaxation methods) we have retained the physical standpoint in this first attack on the case of specified normal gradients. Satisfactory accuracy is attainable.

Method 2 (*for mixed boundary conditions*)

20. An alternative approach to mixed boundary conditions was suggested by J. R. Green. Suppose in the first place that the boundary consists (figure 12) of two portions only, a part ABC on which the wanted function ψ is specified, and a part CDA along which values are imposed upon its normal gradient We assume (§ 1) that ψ being plane-harmonic has a conjugate function ϕ defined by (3), and we arrange the square mesh net so that AC coincides with one line of nodal points

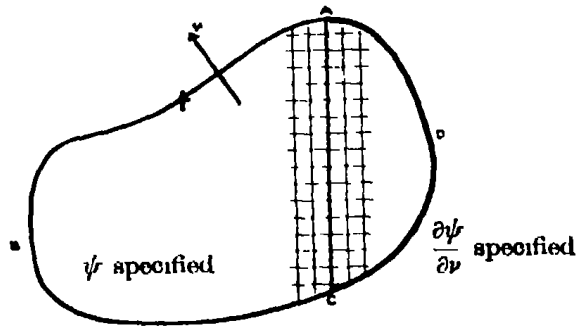


FIGURE 12

Then along CDA, where $\partial\psi/\partial\nu$ is specified, we have values of $\partial\phi/\partial s$ according to the second of (3), and can integrate to obtain ϕ: the constant of integration is immaterial, so ϕ_C can be given any arbitrary value and a definite value will then be imposed upon ϕ_A. We also know ψ_A and ψ_C, but we do not know (initially) the values of ϕ or of ψ at nodal points *between* A and C. If these were given, we should be confronted with two plane-harmonic problems of the first class (§ 1)—namely, evaluation of ψ within the region $ABCA$, and evaluation of ϕ within the region $CDAC$ of figure 12—both soluble by the methods of Part III. Our problem is to attach such values to ϕ and ψ, at nodal points between A and C, that the resulting solutions are compatible; i.e. that ψ as deduced for the region $ABCA$, and ψ' the conjugate of ϕ as deduced for the region $CDAC$, merge without discontinuity along the junction line AC.

21. Identifying the directions of Oy and of CA, we may state the conditions of such mergence as follows:

$$\text{along } AC, \quad \frac{\partial\psi}{\partial x} = \frac{\partial\psi'}{\partial x} = -\frac{\partial\phi}{\partial y}, \text{ by (3)}, \quad \frac{\partial\psi}{\partial y} = \frac{\partial\psi'}{\partial y} = \frac{\partial\phi}{\partial x}, \text{ by (3)}. \tag{28}$$

Let O (figure 13) be any nodal point in the line AC, and let 1, 2, 3, 4 denote the four surrounding points. Then the usual finite-difference approximations replace

$$2a\frac{\partial\psi}{\partial x} \text{ by } \psi_1-\psi_3, \quad 2a\frac{\partial\psi}{\partial y} \text{ by } \psi_4-\psi_2, \tag{i}$$

and similar expressions hold for $\partial\phi/\partial x$, $\partial\phi/\partial y$. But 1, in relation to ψ, is a 'fictitious point', and ψ_1 (since ψ is plane-harmonic) must satisfy the relation

$$\psi_1+\psi_2+\psi_3+\psi_4 = 4\psi_0 \tag{29}$$

Consequently the first of (i) can be replaced by

$$2a\frac{\partial\psi}{\partial x} = 4\psi_0-(\psi_2+2\psi_3+\psi_4), \tag{ii}$$

which involves only values of ψ at 'actual' points in the region $ABCA$ of figure 12. The corresponding approximation to $\partial\phi/\partial x$ is

$$2a\frac{\partial\phi}{\partial x} = 4\phi_0-(\phi_4+2\phi_1+\phi_2), \tag{iii}$$

involving only values of ϕ at 'actual' points in the region $CDAC$ of figure 12.

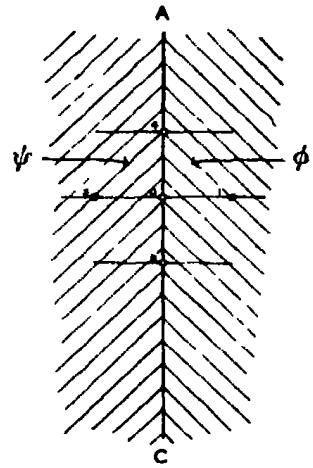


FIGURE 13

The expressions for $\partial\psi/\partial y$, $\partial\phi/\partial y$ call for no modification Substituting from (i), (ii) and (iii), we deduce that the following may be employed as finite-difference approximations to (28).

$$\textit{along } AC, \quad 4\psi_0-(\psi_2+2\psi_3+\psi_4) = \phi_2-\phi_4, \quad 4\phi_0-(\phi_4+2\phi_1+\phi_2) = \psi_4-\psi_2 \tag{30}$$

The mesh-side (a) does not appear in these expressions, which accordingly hold for every size of net, and whether x, y, ϕ, ψ have 'dimensional' or purely numerical significance.

22. We have now transformed our problem into one suited to attack by the standard methods of Part III. The function ψ is to be evaluated at nodal points in the region $ABCA$, ϕ at nodal points in the region $CDAC$, both ψ and ϕ at nodal points in the line AC. The wanted functions are specified on the curved boundary, so the methods of Part III, §§ 23-4, can be employed to deal with 'irregular stars'. Separating the regions $ABCA$ and $CDAC$ as shown in figure 14, we need not (in computation) distinguish between ψ and ϕ. Each is plane-harmonic in its own domain, so special relaxation patterns are entailed only near the curved boundaries (due to irregular stars) and at points near the lines AC, $A'C'$ in figure 14.

For the latter, they may be derived from (30) in the same way that the standard pattern is derived from (29) Thus for the first of (30), introducing the concept of residual forces, we may substitute

$$\mathbf{F}_0 = -4\psi_0+(\psi_2+2\psi_3+\psi_4)+\phi_{2'}-\phi_{4'} = 0 \tag{31}$$

(dashes denoting points on the line $A'C'$ of figure 14). This shows at once the change induced in the residual force at O by unit displacements imposed (severally) at the nodal points numbered 0, 2, 3, 4, 2′, 4′; i.e., it decides one point in the 'relaxation pattern' corresponding with each of these points. Excepting (possibly) at A and C, all of the 'strings' have standard lengths, so the patterns are simple to construct and apply.

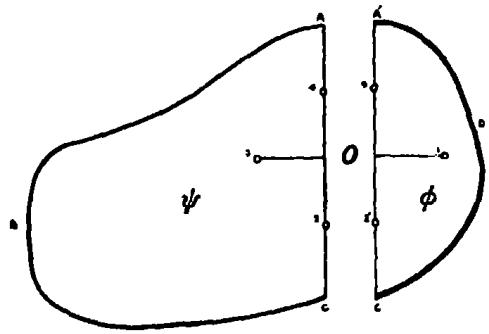


FIGURE 14

In special cases that part of the boundary on which normal gradients are specified may be straight,—i.e. CDA, figure 12, may degenerate into the line CA. Then ϕ will be known initially for points on AC, so that ϕ_2, ϕ_4 are invariant and (31) is recognizable as an expression for $\partial\psi/\partial x$, at O, in terms of ψ_0, ψ_2, ψ_3, ψ_4: no further condition is imposed by the second of (31), since 1, in this instance, is a fictitious point in relation to ϕ as well as ψ, and ϕ_1 is accordingly unrestricted.

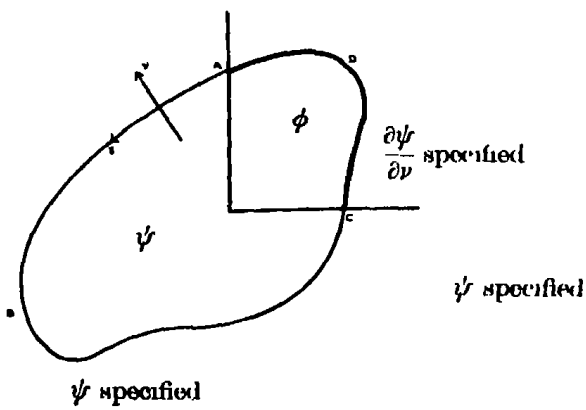


FIGURE 15a

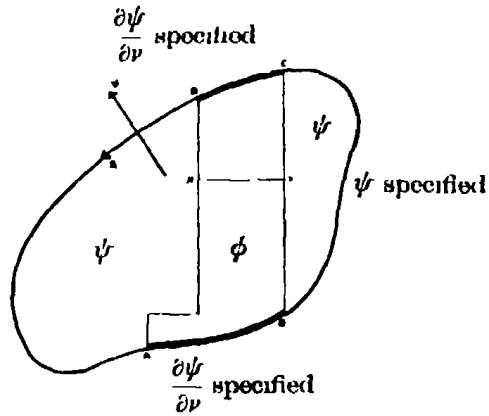


FIGURE 15b

23. Sometimes it may not be convenient to arrange the chosen net (as in § 20) so that AC coincides with a line of nodal points. in that event the barrier between the ψ and ϕ regions may be chosen to coincide with *two* such lines, as in figure 15*a*. No new question of principle is entailed

But our discussion as it stands does not cover problems in which normal gradients are specified on more than one portion of the boundary The reason is indicated in

figure 15*b*: knowing $\partial\phi/\partial s$ at all points in AB, CD, we can attach an arbitrary value to ϕ_A and thence deduce ϕ_B, but we cannot go on to deduce ϕ_C and ϕ_D severally (though we can calculate $\phi_D - \phi_C$), and we are not entitled arbitrarily to introduce another constant in order to make them definite.

We can on the other hand (taking advantage of the principle of superposition) obtain the required solution by a synthesis of two solutions (*a*) and (*b*) of which

(*a*) is obtained on the assumption $\phi_A = \phi_C = 0$,

(*b*) is obtained on the assumption $\phi_A = 0, \phi_C = 1$,

and both can be obtained in the manner of §§ 20–1. Denoting these two solutions by ψ_a, ϕ_a and ψ_b, ϕ_b, we can derive a third solution ψ, ϕ by attaching any value to k in the expressions

$$\psi = \psi_a - k\psi_b, \quad \phi = \phi_a - k\phi_b. \tag{32}$$

To make the derived solution acceptable, k must be so chosen that on any line XY parallel to Ox

$$\psi_Y - \psi_X = \int_X^Y \frac{\partial\psi}{\partial x} dx \quad \text{as determined from } \phi \text{ according to (3)}$$

$$= -\int_X^Y \frac{\partial\phi}{\partial y} dx,$$

so we have from (32)

$$k\{(\psi_b)_Y - (\psi_b)_X\} + \int_X^Y \frac{\partial\phi_b}{\partial y} dx = (\psi_a)_Y - (\psi_a)_X + \int_X^Y \frac{\partial\phi_a}{\partial y} dx. \tag{33}$$

Every line of nodal points like XY will yield a separate estimate of k, and the estimates will differ only on account of unavoidable errors of computation.

Examples 2 and 3

24 To test these methods L Fox applied them to the two problems indicated in figure 16 For both the boundary conditions were framed so that the wanted function was*

$$\psi = \tan^{-1} y/x \quad \text{(corresponding with } \phi = \log r + \text{const.}), \tag{34}$$

when the axes are as shown in the diagrams. But in Example 3 (illustrative of §§ 20–2) boundary values of ψ were assumed to be specified over the length $EBADCH$, normal gradients of ψ over the length $EFGH$, i e a single barrier EH separated a ψ-region from a ϕ-region in Example 4 (illustrative of § 23) boundary values of ψ were assumed to be specified both over $BADC$ and over $EFGH$, normal gradients of ψ over the lengths BE and HC, so that a single ϕ-region separated two ψ-regions. The choice of a rectangular boundary obviated the inessential complication of 'irregular stars'.

Example 3 presented no difficulty. Three successive nets were employed, the

* The advantages of this function have been stated in § 14.

finest having eight mesh-sides in a shorter side of the rectangle. The final (accepted) values of ψ (shown in figure 17) were nowhere in error by $\frac{1}{3}$ %.

The test imposed in Example 4 was somewhat different, being concerned with the accuracy attainable by the procedure suggested in § 23. By that method we derive a series of estimates for the constant k in (32), in number depending on the fineness of the chosen net. In our example we arranged that the correct value of k should be 805, and on a net having eight mesh-sides in the length of a smaller side of the rectangle we obtained as estimated values

$$795,\quad 797,\quad 807,\quad 813,\quad 807,\quad 799,\quad 799 \quad \text{(mean 802)}. \tag{35}$$

Acceptance of the mean value 802 entails an error of less than 0·4 %, which is satisfactory accuracy.

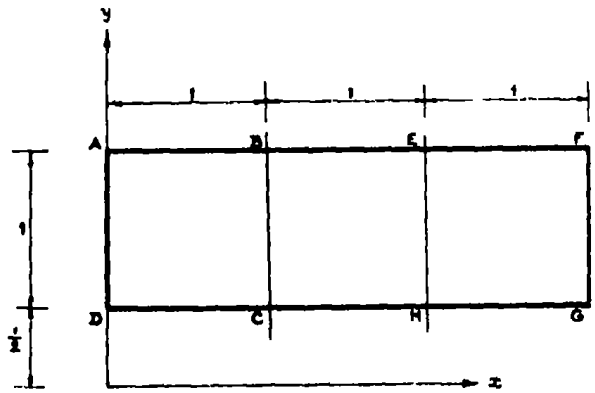


FIGURE 16

CONCLUSION

25. In plane-potential problems, specified normal gradients, whether for the whole or for part of the boundary ('mixed boundary conditions'), inevitably present greater difficulties, and entail more labour for corresponding accuracy, than specified boundary values. But both of the methods described in this paper have been shown by test examples to be adequate, used in conjunction with a reasonably fine net. As in Part III (§ 5), it will almost always be possible to dispose of singularities in advance.

Grateful acknowledgement is made of a grant provided by the Department of Scientific and Industrial Research whereby one of us (G. V.) was enabled to take part in this investigation.

REFERENCES

Bradfield, K. N. E., Hooker, S. G. & Southwell, R. V. 1937 *Proc. Roy. Soc.* A, **159**, 315–346.
Christopherson, D. G. & Southwell, R. V. 1938 (Part III) *Proc. Roy. Soc.* A, **168**, 317–350.
Gandy, R. W. G. & Southwell, R. V. 1940 (Part V) *Phil. Trans.* A, **238**, 453–475.
Shaw, F. S. & Southwell, R. V. 1941 (Part VII) *Proc. Roy. Soc.* A, **178**, 1–17.

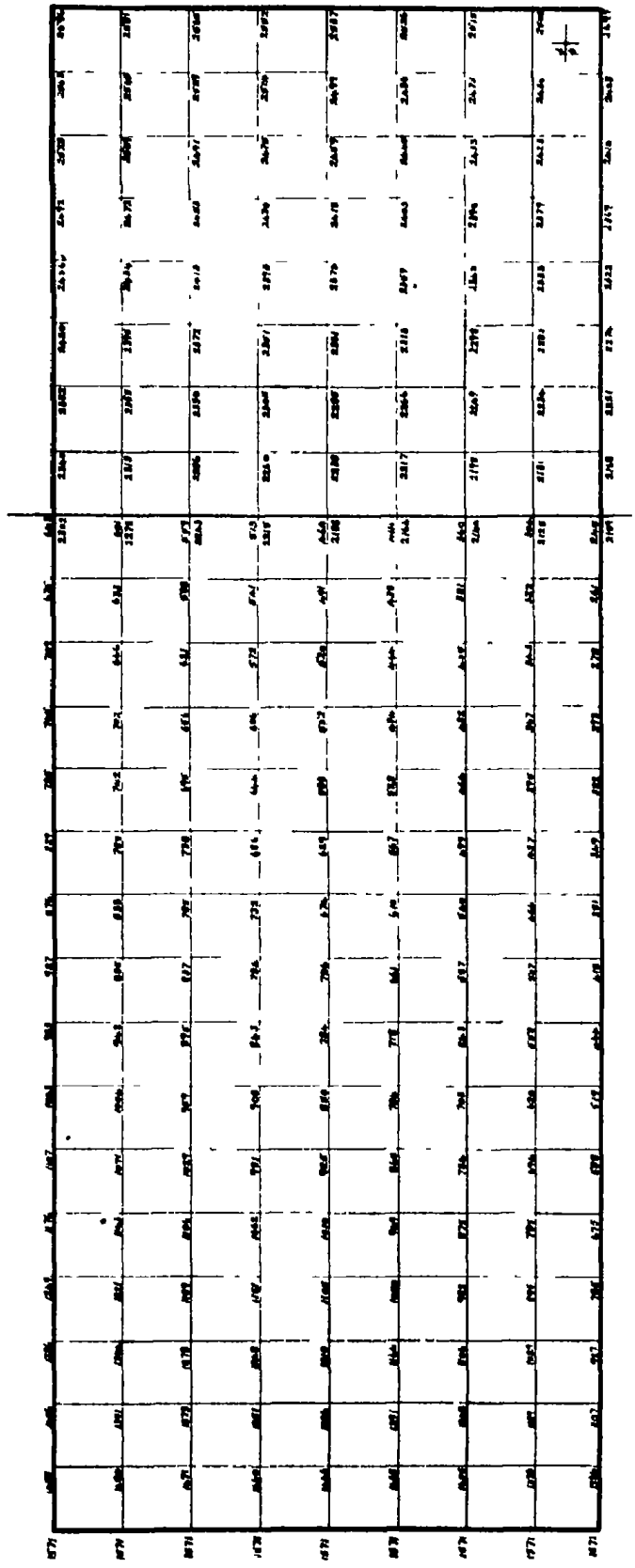


FIGURE 17

The refractive index of an ionized medium. II

By Sir Charles Darwin, F R S.

(*Received* 27 *April* 1943)

The aim of the work is to confirm and to clarify a result in an earlier paper which showed that the refractive index of an ionized medium is given by $\mu^2-1=-4\pi Ne^2/m\nu^2$ and not by the corresponding formula in $3(\mu^2-1)/(\mu^2+2)$. This is done by a direct analysis of the effects of collisions between free electrons and positive ions. It is shown that the perturbation of the electron's path by the light has a secondary effect during a collision, so that if before entry into collision its position was displaced sideways, it will emerge from the collision with a changed velocity. The average effect of such changed velocities is equivalent to an acceleration, which reduces the effective force on the electron from $E+\frac{4\pi}{3}P$ to E. This result is proved first for a positive ion composed of a uniform spherical charge, then for a proton and finally for any centrally symmetrical distribution of charge. Though it has not been proved for a system of unsymmetrical charges arbitrarily orientated, such as ionized molecules, there can be little doubt that the result is general.

1 Introduction

In 1934 I wrote a paper* under this title the aim of which was to show that whereas some writers favoured the use of the Lorentz formula in determining the refractive index of an ionized medium, it was in fact correct to make use of the older Sellmeyer formula. The two formulae are

$$S=\mu^2-1, \tag{1·1}$$

$$L=\frac{3(\mu^2-1)}{\mu^2+2}. \tag{1·2}$$

For most cases the difference between them is not large, so that L may be regarded as a correction to S, and to discuss the distinction may appear somewhat trivial. But it is not trivial for the most interesting case, that is the reflexion of wireless waves from the ionosphere. The question here is whether it is S or L that is to be equated to $-4\pi Ne^2/m\nu^2$, where N is the electron density and ν the radian frequency.† The reflexion is determined by the condition $\mu=0$, and it will be seen that if S is used N is $m\nu^2/4\pi e^2$, but if L it is half as great again. Thus until the matter is settled there is a possibility of a 50 % error in our determination of the number of free electrons in the upper air.

The earlier paper fell into two parts. Its first six sections were a critique of the older methods and aimed at showing how unreliable such methods were. Only negative conclusions were reached, and to overcome the difficulties a new method was introduced in the later sections which avoided them. This method justified the use of the S formula for an ionized medium, and while no claim could be made for

* *Proc. Roy. Soc.* A, **146**, 17 (1934).

† This is 2π times the cyclic frequency. The word frequency will imply radian frequency throughout the paper.

the sort of rigour that appeals to the pure mathematician, the demonstration was at least as good as many that have to be accepted in the more difficult branches of physics. The present note returns to the point of view of the first half of the paper, and shows that from such a starting point the same result is obtained. The previous criticism, which was negative, in that it only pointed out fallacies in the argument favouring L, is now turned into a positive confirmation for S The ideas behind the present work are largely contained in §§ 4–6 of the earlier one

2 Polarization and field strength

In the ordinary development of the theory of refraction the two chief quantities are E the electric field strength and P the polarization Once these are related together the refractive index is given by

$$\mu^2 E = E + 4\pi P \tag{2·1}$$

For an isotropic dielectric medium Lorentz shows that it is reasonable to suppose that the average force acting on the electrons in an atom from outside is

$$F = E + \frac{4\pi}{3} P, \tag{2·2}$$

and this value is justified by comparison of the index of a liquid and its vapour for many, but not all substances. There are many difficulties in taking over these ideas for a medium composed of free electrons interspersed between positively charged atoms: indeed, it was these difficulties of definition which encouraged me in the earlier paper to develop a method entirely free from them Here, as I am returning to the more ordinary method, some consideration must be given to the meanings of P and E

We shall take as our medium a collection of arbitrarily placed fixed positive charges, in the simplest case protons, and moving among them an equivalent number of free electrons at high velocities As it stands the idea of polarization is not present in this, but it is obviously to be introduced by the following considerations. When there is no light present the electrons describe orbits composed of a succession of straight lines in free space and curves past the ions, and if we omit the question of recombination there is no consequent emission of light. When a source of light is superposed it perturbs the electrons' orbits, making the linear parts slightly sinuous and affecting the hyperbolic parts in a manner which we shall have to examine. This motion is described mathematically by saying that the perturbed electron may be replaced by the unperturbed one together with a small dipole travelling with it, and that it is only this dipole which gives the polarization and so is responsible for optical effects.

The field strength E is a much more difficult matter. In the static case of a non-conducting medium composed of polarizable atoms E is defined with the help of a pipe-shaped cavity, whereas F is measured in a spherical cavity. It is sometimes

loosely said that to use F is to allow for the polarization, but this must not be allowed to obscure the point that E as well as F depends on the polarization. For a non-conducting medium these ideas may be taken over without difficulty when E varies with the time as in a light wave, but the matter is very different for a conductor. Here there is no static case at all, and though we have seen that polarization can be given a meaning, it is by no means elementary to see how to define E In particular if the definition is to be made by means of a cavity, some rule must be laid down about what happens when a free electron crosses into this cavity. It has indeed been these difficulties that have caused the doubt attaching to the whole subject

In § 5 of the former paper a derivation was given of the relation of F to E in a dielectric, which is rather different from the ordinary one It was supposed that F was the average force on an electron placed anywhere outside the atoms of a polarized medium. If an electron is carried down the field along an arbitrary line, this line will cut through some of the atoms. During its passage through an atom the electron will come under an opposing force (like the opposing force inside a permanent magnet), and allowance must be made for these 'depolarizing' forces The consequent average force for the whole line is $F - \frac{4\pi}{3} P$, and this at once shows that it is E rather than F that should be regarded as effective on any particle that is not debarred from entering the atoms. For a medium composed of free electrons and fixed charges and no dipoles it is not obvious that there will be a similar effect, but it will be shown by the present work that there is The result is a dynamic not a static one, and it is not possible to deduce it without considering the deflexions of the electrons during their collisions. Consequently no simple consideration of energy will suffice as it does in the static case.

It is possible that satisfactory definitions and reasoning about E could be found which would yield the result directly, but part of the purpose of the present work is to gain insight into the ordinary type of argument. For this purpose I shall be content to adopt Lorentz's argument and shall assume without further discussion that the force acting on an electron is on the average $E + \frac{4\pi}{3} P$ Sufficient justification *in the present argument* is that it was the use of this formula that suggested the correctness of the L expression, whereas it will be shown that on the contrary it leads to S

3. Perturbation effect of collisions

That the collisions may have an important effect is easily seen. If the time of a collision is short compared to the period of the incident light, then during the collision the direct effect of the external force is negligible In figure 1 are constructed a number of sketches of orbits of electrons colliding with a proton. In each diagram there is an unperturbed orbit travelling from A to B, and alongside it a perturbed orbit from A' to B'. Both orbits are simple hyperbolas, only differing in

their starting points. The external force F is down the page, and A' is a constant height ξ above A.

In the line orbit 1 the perturbed electron follows the path of the unperturbed, but a little behind and it is clear that the effect is exactly to reverse the moment This orbit was described in § 6 of the former paper In 2 the unperturbed orbit is again a line, but the perturbed electron has now a small angular momentum and not only is BB' now negative, but the direction of the perturbed path is different so that BB' is increasing negatively with the time. Orbit 3 is similar but with less marked character than 2, again giving BB' negative. Orbit 4 on the other hand shows that the final moment is not necessarily negative, since in this case BB' is both positive and increasing positively as the point B moves away from the centre.

These sketches show that the effects of collisions are very various, and that it is necessary to examine them thoroughly.

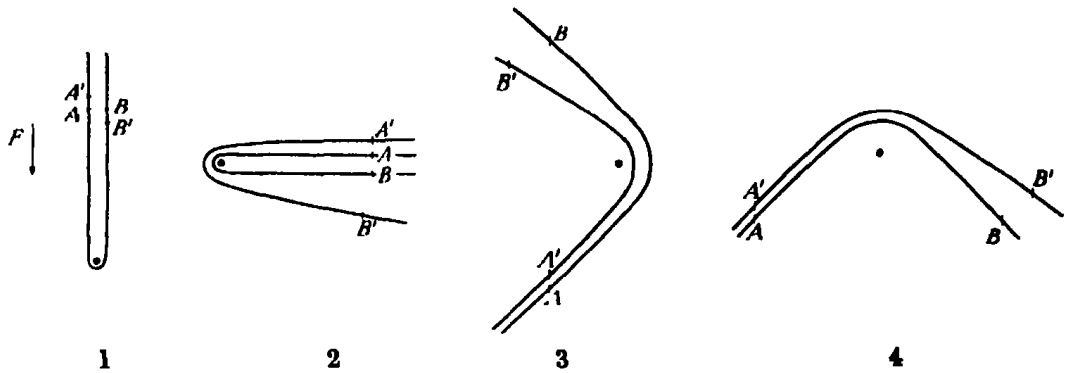


FIGURE 1. Perturbed and unperturbed electrons passing a proton

4. CONDITIONS OF COLLISION

We shall make certain assumptions, satisfied in the important practical cases, which enable us to discriminate sufficiently between the regions where collisions occur and the regions of free space between. At the end a limiting process will remove these assumptions. They are precisely those made in § 9 of the former paper

It is assumed that a sphere of radius b can be described surrounding each positive ion which satisfies the following inequalities:

$$\text{(I) } b \leqslant N^{-\frac{1}{3}}, \quad \text{(II) } b \leqslant V/\nu, \quad \text{(III) } b \geqslant e^2/mV^2.$$

Here N is the numerical density of electrons,

V is the velocity of any electron during its free path,

ν is the frequency of the light,

e and m are charge and mass of the electron.

Condition (I) implies that the regions of collision occupy a very small fraction of space. This makes it improbable that two electrons will be simultaneously in collision with one ion, and the possibility of such double collisions will be neglected.

Condition (II) implies that during a collision the direct effect of the external force is negligible. Since the time of collision is of order b/V, the displacement directly due to F in this time is of order $\frac{eF}{m}\left(\frac{b}{V}\right)^2$, which is much less than the actual displacement which is of order eF/mv^2. Thus in calculating the change of electric moment during a collision the direct effect of F may be neglected, which much simplifies the problem.

Condition (III) implies that at entry into a sphere the electron has only infinitesimally greater velocity than at infinity, because $-e^2/b$ is its potential energy at this point It is thus the natural description for a collision region.

By these conditions we have separated space into the regions of collision and those outside, and may take it that outside the electron is only perturbed by F, the force of the light, while inside any b sphere it is only acted on by the central force of that ion

In addition to colliding with ions the electrons also collide with one another, but it is easy to see that such collisions produce no effect at all In such a collision, though the velocities of the electrons may change, the law of momentum ensures that the sum of the velocities of the pair is unchanged This applies both to the unperturbed and to the perturbed orbits, and therefore to their difference. A collision between two electrons thus produces no change in the sum of their dipole moments, and therefore in the polarization.

5. Uniformly distributed charge

We now have the following problem in central orbits. The unperturbed electron enters the sphere of radius b on an arbitrary line, describes its orbit (for the case of a proton a hyperbola) and emerges. At its moment of entry the perturbed electron is at a point a small distance ξ from it in the direction of F, and is moving with a velocity of which the component in this direction differs by an amount $\dot{\xi}$ At a certain later time the unperturbed electron emerges from the sphere What is the relative position and velocity of the perturbed electron at this moment? Since transverse displacements will average out, it is only the component in the direction of F that matters.

The calculations are fairly intricate even for the case of the hyperbola described round a proton, and it will make the whole argument more evident if we first discuss a much simpler model for which the solution is elementary This consists in a sphere of radius b entirely filled with a uniform distribution of positive electricity of total charge $-e$.

At the surface the electron is attracted with force e^2/b^2. To simplify writing we shall set

$$e^2/m = M, \tag{5·1}$$

and shall regard the electron as having unit mass. At any point inside the sphere a component of the force is then $-Mx/b^3$.

If $M/b^3 = k^2$, the equations of motion are

$$x + k^2 x = 0. \tag{5·2}$$

Let l be the direction cosine of approach for the unperturbed electron, and let λ be the direction cosine of the radius drawn towards the centre from the point of entrance. The orbit is then

$$x = -\lambda b \cos kt + (lV/k) \sin kt \tag{5·3}$$

The time of exit is easily found to be given by

$$\tan kt_1 = 2bVk(\lambda l)/(V^2 - b^2k^2).$$

There is no need to carry out the exact solution, because in view of condition (III) $V^2 \geqslant M/b = b^2k^2$, so that kt_1 is a small angle and we may write

$$t_1 = 2b(\lambda l)/V. \tag{5·4}$$

This merely means that on account of the high velocity the orbit is nearly straight and the time is practically that which would be taken in moving on a straight line at constant speed along a chord of the sphere.

The perturbed electron starts at the same time but from position differing by ξ and with velocity differing by $\dot{\xi}$, in the direction of F, which we will take as the z-direction. Then the perturbation at any time is

$$\delta z = \xi \cos kt + (\dot{\xi}/k) \sin kt,$$

and approximating sufficiently we have at exit

$$\delta z = \xi + \dot{\xi} 2b(\lambda l)/V, \tag{5·5}$$

$$\delta \dot{z} = \dot{\xi} - \xi 2M(\lambda l)/b^2V. \tag{5·6}$$

The new term in (5·5) is unimportant since it does not depend on M and merely signifies that ξ grows steadily in the time t_1 on account of the initial value of $\dot{\xi}$. The important effect of the collision is that, as shown in (5·6), the initial displacement is responsible for a final velocity. Figure 1 illustrates this fact. Thus the collision produces a change of velocity of amount

$$-\xi 2M(\lambda l)/b^2V.$$

We are to find the average consequences of this. Taking the average for all directions and lines of approach, we have

$$\overline{(\lambda l)} = \tfrac{2}{3}.$$

During a time Δt the chance of a given electron colliding with any of the positive charges is

$$N\pi b^2 V \Delta t, \tag{5·7}$$

and so the average effect of collisions is that in each interval of time Δt the velocity is changed by an amount

$$\Delta \dot{\xi} = -\xi \frac{4\pi}{3} NM\Delta t. \tag{5·8}$$

The important point about this is to note that b has cancelled out, so that the artificial assumption of a separation between collision and non-collision regions has disappeared Observe also that V has disappeared so that the result in no way depends on the distribution of electron velocities, apart from the condition that V must be large.

The effect of collisions on an electron may be re-expressed by saying that on the average they produce an acceleration

$$-\frac{4\pi}{3}NM\xi.$$

Replacing M from (5·1) the equation of motion of the electron is therefore

$$\ddot{\xi} = \frac{eF}{m} - \frac{4\pi}{3}\frac{Ne^2}{m}\xi \tag{5·9}$$

Now $$Ne\xi = P \quad \text{and} \quad F = E + \frac{4\pi}{3}P,$$

so that this gives $\xi = eE/m$ and therefore $P = -Ne^2E/m\nu^2$,

from which by (2·1) $$\mu^2 - 1 = -4\pi Ne^2/m\nu^2,$$

that is the N formula.

6 General formula for effect

The preceding section discussed a very artificial model and we have to evaluate the quantity similar to (5·8) for other more natural atoms. It will be shown that exactly the same result follows when the positive charges consist in any distribution of electricity with perfect spherical symmetry. The method of general dynamics is the most convenient method of attack on the problem

Since the mass of the electron is taken as unity, the co-ordinates and velocities $x, y, z; \dot{x}, \dot{y}, \dot{z}$ constitute a canonical set of variables. We apply a canonical transformation into the following variables P_1 the energy, P_2 the angular momentum in the plane of the orbit, P_3 the angular momentum about the z-axis which is the direction of the initial displacement ξ. P_2 is to be taken positive, and the inclination of the orbit is defined by $\cos i = P_3/P_2$. The conjugate variables are Q_1, the time measured from the apse, Q_2 the angle from the ascending node to the apse, and Q_3 the longitude of the ascending node.* In the plane of the orbit we take a pair of axes of which X is along the apse and Y at right angles Then X and Y, $\dot{X}$ and $\dot{Y}$ are functions of P_1, P_2 and Q_1 in a manner depending on the law of force, but are independent of the other three variables.

* A hyperbolic orbit itself may have only a single node, so that the term 'ascending node' is sometimes a misdescription It is not hard, however, to see that a complete and unique description of all orbits is given by considering both intersections of the plane of the orbit with the xy plane as nodes. The complete set of orbits is then included in the ranges

$$0 < i < \pi, \quad 0 < Q_2 < 2\pi, \quad -\pi < Q_3 < \pi.$$

The canonical transformation converts the unperturbed orbit into $Q_1 = t - t_a$, where t_a is the time of passage through the apse; the other five elements are all constant At entry $Q_1^0 = -t_a$, and the time of exit is $2t_a$. For the perturbed orbit the five constant elements are all slightly changed and the time is given by

$$Q_1 = t - t_a - \delta t_a.$$

The entry and exit times will be counted as those of the unperturbed orbit regardless of the fact that the perturbed electron is not at those moments exactly on the sphere of radius b. Then in Q_1-space the perturbed path is parallel to the unperturbed at a distance $\delta Q_1 = -\delta t_a$ above it, so that δQ_1 is a constant of the motion irrespective of the fact that Q_1 is the current co-ordinate.

It was seen in § 5 that the important quantity to evaluate was the ξ at exit arising from ξ at entry, and that the ξ at exit did not matter, nor did the exit values consequent on ξ at entry. The same result must be established in general, but it will be convenient to defer the discussion until § 9. We have therefore to determine the variations in the six elements at entry due to the perturbation

$$\delta x = \delta y = 0, \quad \delta z = \xi, \quad \delta \dot{x} = \delta y = \delta \dot{z} = 0.$$

For P_1 this is evidently $$\delta P_1 = \xi \left(\frac{\partial P_1}{\partial z}\right)_0.$$

It is to be evaluated at the point of entry. By the property of canonical transformations this may be replaced by

$$\delta P_1 = -\xi \left(\frac{\partial \dot{z}}{\partial Q_1}\right)_0.$$

Similarly we have $$\delta Q_1 = \xi \left(\frac{\partial \dot{z}}{\partial P_1}\right)_0$$

and similar relations for the other variables.

Now $$\dot{z} = \sin i (\dot{X} \sin Q_2 + \dot{Y} \cos Q_2), \tag{6·1}$$

so that we have

$$\left.\begin{aligned}
\delta P_1 &= -\xi \sin i \left[\frac{\partial \dot{X}}{\partial Q_1} \sin Q_2 + \frac{\partial \dot{Y}}{\partial Q_1} \cos Q_2\right]_0, \\
\delta Q_1 &= \xi \sin i \left[\frac{\partial \dot{X}}{\partial P_1} \sin Q_2 + \frac{\partial \dot{Y}}{\partial P_1} \cos Q_2\right]_0, \\
\delta P_2 &= -\xi \sin i \, [\dot{X} \cos Q_2 - \dot{Y} \sin Q_2]_0, \\
\delta Q_2 &= \xi \left[\sin i \left(\frac{\partial \dot{X}}{\partial P_2} \sin Q_2 + \frac{\partial \dot{Y}}{\partial P_2} \cos Q_2\right) + \frac{\cos^2 i}{P_2 \sin i} (\dot{X} \sin Q_2 + \dot{Y} \cos Q_2)\right]_0, \\
\delta P_3 &= 0, \\
\delta Q_3 &= -\xi \frac{\cos i}{P_2 \sin i} [\dot{X} \sin Q_2 + \dot{Y} \cos Q_2]_0.
\end{aligned}\right\} \tag{6·2}$$

We have to estimate the effect that the quantities (6·2) will produce on $\dot{z}$ at the point of exit, and this is made possible by the symmetry of a central orbit about the apse. Thus instead of expressing δP_1, etc. in terms of the entry point we can express them in terms of the exit point. To do this we note that this point has $Q_1 = -Q_1^0$ and that

$$X,\ \dot{Y},\ \frac{\partial \dot{X}}{\partial \bar{Q}_1},\ \frac{\partial \dot{Y}}{\partial \bar{P}_1},\ \frac{\partial \dot{Y}}{\partial \bar{P}_2} \quad \text{are even functions of } Q_1$$

while

$$Y,\ \dot{X},\ \frac{\partial \dot{Y}}{\partial \bar{Q}_1},\ \frac{\partial \dot{X}}{\partial \bar{P}_1},\ \frac{\partial \dot{X}}{\partial \bar{P}_2} \quad \text{are odd functions.}$$

Then, denoting the exit point by X, etc., without affixes

$$\left.\begin{aligned}
\delta P_1 &= -\xi \sin i \left[\frac{\partial \dot{X}}{\partial \bar{Q}_1} \sin Q_2 - \frac{\partial \dot{Y}}{\partial \bar{Q}_1} \cos Q_2\right],\\
\delta Q_1 &= \xi \sin i \left[-\frac{\partial \dot{X}}{\partial \bar{P}_1} \sin Q_2 + \frac{\partial \dot{Y}}{\partial \bar{P}_1} \cos Q_2\right],\\
\delta P_2 &= -\xi \sin i \left[-\dot{X} \cos Q_2 - \dot{Y} \sin Q_2\right],\\
\delta Q_2 &= \xi \left[\sin i \left(-\frac{\partial \dot{X}}{\partial \bar{P}_2} \sin Q_2 + \frac{\partial \dot{Y}}{\partial P_2} \cos Q_2\right) + \frac{\cos^2 i}{P_2 \sin i}(-\dot{X} \sin Q_2 + \dot{Y} \cos Q_2)\right]
\end{aligned}\right\} \tag{6·3}$$

The consequent change in $\dot{z}$ is given by

$$\delta \dot{z} = \left(\frac{\partial z}{\partial \bar{P}_1}\right) \delta P_1 + \quad .$$

Writing in $\dot{z}$ from (6·1) and substituting from (6·3) the formula can be considerably reduced. After the reduction one further simplification can be made, for

$$\frac{\partial X}{\partial Q_1} = X = -M \frac{X}{b^3},$$

which is useful since b is to become large.

The result is

$$\begin{aligned}
\delta \dot{z} = {} & \frac{2\xi \cos^2 i}{P_2} \dot{X} \dot{Y} + \xi \sin^2 i \left[M \frac{X}{b^3} \frac{\partial \dot{X}}{\partial \bar{P}_1} - M \frac{Y}{b^3} \frac{\partial \dot{Y}}{\partial P_1} + \frac{\partial}{\partial \bar{P}_2}(\dot{X} \dot{Y})\right]\\
& + \xi \sin^2 i \cos 2Q_2 \left[-M \frac{X}{b^3} \frac{\partial \dot{X}}{\partial \bar{P}_1} - M \frac{Y}{b^3} \frac{\partial \dot{Y}}{\partial P_1} + \dot{X} \frac{\partial \dot{Y}}{\partial \bar{P}_2} - \dot{Y} \frac{\partial \dot{X}}{\partial \bar{P}_2}\right].
\end{aligned} \tag{6·4}$$

This is now to be averaged for all directions and lines of approach, that is to say for P_2, P_3, Q_2, Q_3. It will appear that it is unnecessary to average for P_1. The averages for Q_2, Q_3 are obvious, for the others we have $\int dP_2 dP_3 = \int P_2 dP_2 \sin i\, di$. The average

of $\cos^2 i$ is 1/3 and of $\sin^2 i$ is 2/3. Moreover, the greatest value of P_2 is approximately bV, so that the normalizing factor for P_2 is $2/b^2V^2$ and we have

$$\Delta\xi = N\pi b^2 V\Delta t \frac{2}{b^2V^2}\int_0^{bV} P_2 dP_2 \left\{\frac{1}{3}\frac{2\xi\dot{X}\dot{Y}}{P_2}+\frac{2}{3}\xi\left[M\frac{X}{b^3}\frac{\partial\dot{X}}{\partial P_1}-M\frac{Y}{b^3}\frac{\partial\dot{Y}}{\partial P_1}+\frac{\partial}{\partial P_2}(\dot{X}\dot{Y})\right]\right\}$$

or
$$\Delta\xi = \xi\frac{4\pi}{3}\frac{N\Delta t}{V}\int_0^{bV}\left\{\frac{\partial}{\partial P_2}(P_2\dot{X}\dot{Y})+\frac{MP_2}{b^3}\left(X\frac{\partial\dot{X}}{\partial P_1}-Y\frac{\partial\dot{Y}}{\partial P_1}\right)\right\}dP_2. \tag{6·5}$$

It may be noted that the integrand contains Q_1 implicitly, so that the first term cannot be integrated as it stands.

7. Application to case of § 5

Before attacking the proton it will be well to exhibit the result for the model of § 5, since there is a rather difficult point in the proton case which will be illustrated thereby. For the uniformly distributed charge of § 5 we have

$$X = f\cos kQ_1, \quad Y = g\sin kQ_1, \tag{7·1}$$

where f is less than b, and g much greater than b. Then, measuring the potential energy from infinity, we have

$$\left.\begin{aligned} P_1 &= \tfrac{1}{2}k^2(f^2+g^2)-\tfrac{3}{2}k^2b^2,\\ P_2 &= kfg,\\ P_2\dot{X}\dot{Y} &= -kP_2^2\sin kQ_1\cos kQ_1,\\ X\frac{\partial\dot{X}}{\partial P_1}-Y\frac{\partial\dot{Y}}{\partial P_1} &= \left(-f\frac{\partial f}{\partial P_1}-g\frac{\partial g}{\partial P_1}\right)k\sin kQ_1\cos kQ_1 = -(1/k)\sin kQ_1\cos kQ_1. \end{aligned}\right\} \tag{7·2}$$

Thus in (6·5) the first term yields

$$-2kP_2\sin kQ_1\cos kQ_1$$

and the second
$$-kP_2\sin kQ_1\cos kQ_1.$$

From these Q_1 is to be eliminated through the condition, expressed in terms of P_1, P_2 and Q_1, that $X^2+Y^2 = b^2$, and the simplest way to do this is to recall that the orbit is nearly a straight line in the Y direction. Then kQ_1 is a small angle, and Q_1 is the time of flight from the apse, that is $\sqrt{(b^2-f^2)}/V$, while $P_2 = fV$ to sufficient approximation. The integral is thus

$$\int_0^{bV} -3kP_2\frac{k\sqrt{(b^2V^2-P_2^2)}}{V^2}dP_2, \tag{7·3}$$

which on substitution for k^2 gives $-MV$, and so verifies (5·8).

8. Application to proton

We now consider the case of a proton. The orbit here is most conveniently described in terms of major axis a, eccentricity $\sec\gamma$, and ψ the hyperbolic eccentric anomaly. In these variables

$$\left.\begin{aligned} X &= a(\sec\gamma - \cosh\psi), \quad Y = a\tan\gamma\sinh\psi, \\ r &= a(\sec\gamma\cosh\psi - 1), \\ \dot{X} &= -\sqrt{(Ma)}\sinh\psi/r, \quad \dot{Y} = \sqrt{(Ma)}\tan\gamma\cosh\psi/r, \end{aligned}\right\} \tag{8·1}$$

while the canonical variables are

$$P_1 = M/2a, \quad P_2 = \sqrt{(Ma)}\tan\gamma, \quad Q_1 = \sqrt{(a^3/M)}\,[\sec\gamma\sinh\psi - \psi]. \tag{8·2}$$

Also
$$V = \sqrt{(M/a)}.$$

The condition (III) implies immediately that $a \leqslant b$ and we shall therefore expand the necessary expressions in powers of a/b. An inspection of the integrand in (6·5) reveals a great distinction between the present case and that of the preceding section, for there the two terms were of similar magnitude, whereas here they are quite different. In the second term X is of order b and P_2 at the outer end is so too, whereas $\dot{X}$ and $\partial\dot{X}/\partial P_1$ are independent of b. There is a factor b^3 in the denominator and therefore it is only necessary to calculate the leading term of its expansion On the other hand the first term of (6·5) has a factor P_2 in some of its parts, and on integration might yield a term in b^2 and therefore it is necessary to carry the expansion for it to three terms.

The expression of the various differential coefficients in terms of the new variables calls for no comment. After carrying it out we substitute $\cosh\psi = \cos\gamma(b+a)/a$, and to simplify writing the formulae shall retain $\tanh\psi$, which is

$$[1 - \sec^2\gamma\, a^2/(b+a)^2]^{\frac{1}{2}},$$

and so is nearly unity except for orbits where γ is nearly a right angle.

To the approximations needed

$$\left.\begin{aligned} X &= -b\cos\gamma, \quad Y = b\sin\gamma\tanh\psi, \\ \frac{\partial\dot{X}}{\partial P_1} &= -\sqrt{\frac{a}{M}}\tanh\psi\cos^3\gamma, \\ \frac{\partial\dot{Y}}{\partial P_1} &= \sqrt{\frac{a}{M}}\sin\gamma(1+\cos^2\gamma) \end{aligned}\right\} \tag{8·3}$$

This term therefore gives in the integrand

$$M\frac{a}{b^2}\tanh\psi\tan\gamma(2\cos^4\gamma - 1). \tag{8·4}$$

For the first term we have

$$\left.\begin{aligned} \hat{X} &= -\sqrt{\frac{M}{a}}\tanh\psi\cos\gamma\left(1+\frac{a}{b}\right), \\ \hat{Y} &= \sqrt{\frac{M}{a}}\sin\gamma\left(1+\frac{a}{b}\right), \\ \frac{\partial\hat{X}}{\partial P_2} &= \frac{\tanh\psi}{a}\sin\gamma\left[\cos^2\gamma\left(1+\frac{a}{b}\right)+\sin^2\gamma\frac{a^2}{b^2}\right], \\ \frac{\partial\hat{Y}}{\partial P_2} &= \frac{\cos\gamma}{a}\left[\cos^2\gamma\left(1+\frac{a}{b}\right)+\sin^2\gamma\frac{a^2}{b^2}\right] \end{aligned}\right\} \tag{8·5}$$

This term yields

$$-\frac{M}{a}\left(1+\frac{a}{b}\right)^2\tanh\psi\tan\gamma\left[2\cos^4\gamma-\frac{a^2}{b^2}(2\cos^4\gamma-3\cos^2\gamma+1)\right]. \tag{8·6}$$

The necessity for this careful evaluation arises from the fact that at the outermost point γ is nearly a right angle so that $\tan\gamma$ becomes large, and also dP_2 gives a factor in $\sec^2\gamma$. Adding the terms together the integral becomes

$$\int_0^{\tan^{-1}b/a}\sqrt{Ma}\sec^2\gamma\,d\gamma\left\{-\frac{M}{a}\left(1+\frac{a}{b}\right)^2\tanh\psi\tan\gamma\,.\,2\cos^4\gamma\right.$$
$$\left.+\frac{Ma}{b^2}\tanh\psi\tan\gamma(4\cos^4\gamma-3\cos^2\gamma)\right\}, \tag{8·7}$$

and it will be observed that the two terms in Ma/b^2 not involving $\cos^2\gamma$ have cancelled. The last term may be integrated and is found to involve b in the form $\log b/b^2$ which is negligible The leading terms of (8·7) may be integrated up to $\pi/2$ without significant error, and the factors $\tanh\psi$ and $(1+a/b)^2$ are unity to this approximation. The integral is thus

$$-\sqrt{(M^3/a)} = -MV,$$

so that

$$\Delta\xi = -\xi\frac{4\pi}{3}NM\Delta t, \tag{8·8}$$

as it was in § 5. The remainder of the argument goes as before and yields the S formula.

The cancellation of the two terms, referred to in the paragraph above, is important, for each yields a term $\frac{1}{3}MV$, which taken by itself would entirely have changed the refraction formula. This cancellation explains why it is possible that in the proton the leading term of (6·5) should be responsible for practically the whole of the effect, whereas for the distributed charge the second term contributed one-third of it. This seemed paradoxical because there is no very different quality in the two fields for orbits which do not pass very near the origin, so that it would have been expected that for the proton also the second term would make an important contribution.

The paradox has been cleared up by seeing that the first term yields a contribution which cancels the effect of the second.

The expressions (8·5) taken only to their leading terms may also be substituted in (6·4) to yield the data from which the curves of figure 1 were drawn. The result is

$$\delta\dot{z} = -(\xi V/a)\cos^2\gamma\{2\cos^2 i + \sin^2 i(\cos 2\gamma + \cos 2Q_2)\}. \tag{8·9}$$

The extreme values of this are respectively $-2\xi V/a$ for the line hyperbola at the equator in figure 1·2 and $+\frac{1}{2}\xi V/a$ for the rectangular hyperbola in figure 1·4.

9. Other effects of perturbation

We must now examine the perturbations due to $\dot{\xi}$ at entry, and verify their unimportance. The calculations follow the same course as in § 6 and § 8. They will not be given, but for the guidance of anyone wishing to check the rather intricate calculations it may be mentioned that in these cases it is the second term of the expansion in powers of a/b that yields the leading term in the result, and that in view of this the greatest value of P_2 must be taken as $(b^2+2ab)^{\frac{1}{2}}V$, instead of the approximate value bV.

The results are:
$$\Delta\dot{\xi} = -\xi\frac{4\pi}{3}NM\Delta t - \dot{\xi}\frac{4\pi}{3}NM\frac{b}{V}\Delta t, \tag{9·1}$$

$$\Delta\xi = -\xi\frac{4\pi}{3}NM\frac{b}{V}\Delta t + \dot{\xi}\frac{4\pi}{3}Nb^3\Delta t. \tag{9·2}$$

The order of magnitude of $\dot{\xi}$ is $\nu\xi$, so that the second term in (9·1) bears ratio $\nu b/V$ to the first, and this is negligible by condition (II) of § 4. The first term of (9·2) can be best understood by reference to figure 1. It will be seen that the perturbed orbit roughly follows the unperturbed up to the apse and then branches off from it in a different direction. The time of flight of this second half is roughly b/V, so that the first term of (9·2) merely expresses the same fact as the first term of (9·1), and since the effect is fully allowed for by (9·1) there is no need to pay attention to it. The second term of (9·2) is negligible on account of condition (I) of § 4. It arises simply from the fact that on the average an electron spends a fraction $\frac{4\pi}{3}Nb^3$ of its time in collision, so that during time Δt the value of ξ is increased on account of collisions by this fraction of Δt multiplied by $\dot{\xi}$.

In consequence of these considerations it will be seen that the first term of (9·1) is the only one that matters, so that the neglect of the rest was justified in §§ 6–8.

10. Extension to general central field

Now that the convergence of the integral has been established, a much simpler method can be used which can be extended to more complicated fields of force. From (8·5) we see that

$$\dot{X} = -\sqrt{(M/a)}\cos\gamma, \quad \dot{Y} = \sqrt{(M/a)}\sin\gamma \tag{10·1}$$

as first approximation. These hold except for orbits in which the apse is far from the centre, where they would break down because their ψ is no longer large. We have seen that these remote orbits give no contribution to the integral, and may therefore omit the corrections they would make The approximations (10·1) are now independent of Q_1 and therefore the first term of (6·5) can be integrated as it stands and gives

$$\left[P_2 \dot{X} \dot{Y}\right]_{\gamma=0}^{\pi/2}, \tag{10·2}$$

which is $-MV$ just as before.

Consider now any spherically symmetrical distribution of charge in the ion, say a nucleus and a cloud of electrons rigidly bound round it If the apse is outside the cloud the orbit will be just as before, but, if the apse is inside the cloud, the near part of the orbit will be changed in a manner that depends on the law of force. At some later point in its orbit however the electron will emerge from the cloud and will then describe the remaining part of a hyperbola If this hyperbola were produced backwards it would come to an apse on a line at some angle α to the true apse. The hyperbola is then described by a and γ as before together with α which, since we do not need to consider variations in the energy, may be regarded as a function of γ. This function could be determined if we were given the field of force, but all that we need to know about it is that it vanishes when the apse is outside the cloud, that is to say when P_2 and therefore γ is greater than a certain quantity

To evaluate the integral we are only concerned with the value of the velocity on the sphere b. This is now

$$\dot{X} = -\sqrt{(M/a)}\cos(\gamma-\alpha), \quad \dot{Y} = \sqrt{(M/a)}\sin(\gamma-\alpha), \tag{10·3}$$

and so the integral is

$$-\sqrt{(M^2/a)}\left[\tan\gamma\cos(\gamma-\alpha)\sin(\gamma-\alpha)\right]_{\gamma=0}^{\pi/2}.$$

Since $\alpha = 0$ at the upper limit, and $\gamma = 0$ at the lower, this again gives $-MV$. Though it has little practical interest, it may be noted that this result is also true when there is no point charge at the centre, even though in that case the orbit of zero momentum passes right across the origin instead of being returned along the way it came.

One further generalization is easy We have hitherto taken the nucleus as singly ionized, but this is unnecessary. Suppose that it is ionized s times. M is altered into sM. In forming (5·7) however N occurred as the number of ions which a given electron would meet, and since N is defined as the electron density we must now substitute N/s in order that the gas should be uncharged. The factors therefore cancel and the degree of ionization is immaterial.

11. Discussion of generalization

It has been shown that for a centrally symmetrical ion of the most general type, the process of collision produces dynamically a 'depolarizing' effect, reducing the

effective average force on an electron from F to E. It is hardly possible to doubt that this result is true for an unsymmetrical distribution of charges in the ion, for example, for an isotropic assembly of ionized molecules, but the present method cannot handle the problem, since it essentially depends on the symmetry of the orbit about the apse.

It is not easy to see how this more general result could be proved, but it may be of some interest to state it as a curious theorem in attractions.

An arbitrary distribution of attracting matter is rigidly held together in a limited space. In gravitational units its total mass is M. A sphere of large radius b is drawn round it. A unit particle approaches this sphere at velocity V, this velocity being so high that $V^2 \geqslant M/b$, describes its orbit and emerges. If this orbit before entry is displaced parallel through a small distance ξ along the z-direction, the orbit will be changed, and the emergent place and velocity with it. Let the z-component of the emergent velocity differ from that for the first orbit by $\delta\dot{z}$. Let this now be averaged for all lines and directions of approach, and also for all orientations of the attracting matter. The theorem then asserts that

$$\delta\dot{z} = -\frac{4}{3}\frac{M}{b^2 V}\xi. \tag{11·1}$$

If such a theorem could be established the proof would be complete. However it is already quite clear what the value of the refractive index must be, and the whole subject perhaps hardly merits much further intensive study.

The electron diffraction by amorphous polymers

By G. D. Coumoulos

(*Communicated by Eric K. Rideal, F.R.S.—Received 2 February* 1943)

[Plate 2]

The configurations of the polyvinyl acetate, and the acrylate and methacrylate polymers revealed by electron photographs suggest a zigzag C-atom chain for the long main chain, which has the 1, 3 structure, with the side-chains alternately on the right and the left of the zigzag chains and on planes approximately perpendicular to the axis of the main chain. These side-chains are subject to lateral cohesive forces, which group them in clusters. In the clusters the side-chains tend to arrange themselves parallel to one another. In the lenses the clusters consist of a small number of side-chains without any apparent arrangement, whilst the multilayer pattern indicates a certain orientation of the side-chains with perhaps a larger number grouping together.

The patterns indicate an 'amorphous' character which is attributed to the tendency of the side-chains to close-packing in clusters, producing distortion of the main chain and thus preventing adlineation.

On the basis of this configuration some of the elastic properties of these polymers are discussed, and a note is made on the occurrence of high elasticity.

Whilst a number of polymeric materials reveal under X-ray or electron diffraction examination characteristic crystalline diagrams, others show such diagrams only when extended under suitable conditions of temperature, and normally in their unextended form present an 'amorphous' pattern, retained in several cases even on stretching.

The X-ray and electron diffraction photographs of polymers revealing these amorphous diagrams usually consist of more or less well-defined haloes on a background which is more dense than that obtaining with crystalline materials.

There exists at present a good deal of uncertainty in the conclusions drawn as to the structure and the molecular distances derived from measurements of these haloes. Early work on the derivatives of cellulose suggested that the sharp rings presented by the electron photographs were due to the presence of numerous imperfect crystallites present in the material or at least due to portions of material in which certain lengths of the macromolecular chains in different molecules were adlineated in respect to one another

Katz (1927, 1936) appears to be the first to have drawn in a tentative way an analogy between the X-ray patterns of amorphous polymers and those of the monomers of the same materials in the liquid state This analogy was not pursued by him in detail, since he noted that in certain polymers a new ring not present in the monomeric material was obtained, these polymeric rings are found at the following Bragg spacings

polymer	spacing in A
styrene	10 0 (Katz 1936)
polyvinyl acetate	7 0 (Katz 1936)
indene	9 1 (Beal, Anderson & Long 1932)
cumarone	9 1 (Beal *et al* 1932)

Katz suggested that this spacing was essentially the distance between parallel chains of the macromolecules separated by a side-chain spacing from one another.

Stewart (1930) suggested that the diffraction effects in liquids were due to cybotaxis Warren (1933) first applied the Zernicke-Prins method to X-ray diffraction by long-chain molecules of organic liquids, expressing the arrangement of the molecules in the liquid by means of a distribution function, the peaks thus giving spacings within one molecule or the most frequently occurring distances between molecules which in relatively densely packed liquids are necessarily dependent on molecular configuration.

Later, Simard & Warren (1936) attempted a detailed analysis of the X-ray photographs of amorphous rubber on the basis of an X-ray study of liquids and concluded that the superficial resemblance of the pattern to that given by a liquid was borne out by the analysis

It seemed possible that further insight into the structure of these synthetic polymeric systems might be obtained by examining the electron diffraction of a series of synthetic polymers of known composition in which the length of the side-chain in the repeating unit was progressively increased. At the same time it seemed

desirable to make a comparison between the diffraction phenomena observed on passage of a beam of electrons through a thin sheet of the polymer and through a multilayer formed from the same polymeric material. In the latter case a certain degree of artificial molecular orientation is imposed. One might anticipate a marked increase in the crystallite content of the material if these do in fact exist in the polymer, whilst if the polymer is to be regarded merely as a liquid, the intensity of certain intermolecular distances might be expected to show a small increase. I have, in fact, observed a difference in the electron-diffraction patterns between the lens and the multilayer and from a comparison one can draw certain conclusions, not only in respect to the orientation existing in the multilayer, but also about the molecular configuration in these films The conclusion arrived at is that these polymers which exhibit amorphous electron-diffraction features in their patterns are in a liquid state.

Experimental

The electron-diffraction camera was that used in a previous investigation (Cameron & Coumoulos 1941) The wave-length of the accelerated electrons was $\lambda = 0{\cdot}06$–$0{\cdot}07$ A, calibration being made by superposition of a pattern of kaolin The following polymers were investigated

polymer	mol. wt. (by viscosity)
vinyl acetate	(*a*) 62,800
	(*b*) 7,100
	(*c*) 26,000
methyl acrylate	7,000
ethyl acrylate	12,000
methyl methacrylate	9,000
ethyl methacrylate	200,000
n-butyl methacrylate	19,000
cyclohexyl methacrylate	—
β-ethoxy-ethyl methacrylate	(*a*) 33,000
	(*b*) 27,000

Dilute solutions of the polymers in acetone, toluene or benzene were placed on the surface of water saturated with the solvent After slow evaporation of the solvent the thin lens (ca. 10^{-5} cm.) was picked up on the specimen holder (Cameron & Coumoulos 1941) and placed in position in the camera for electron transmission. It has been shown that certain polymers can be spread quantitatively to give monolayers at an air-water interface under these conditions. They are subjected to strong orientational forces which do not persist in the bulk of a lens (Crisp 1942), and forming a multilayer by deposition of successive monolayers is the best method of achieving the most favourable conditions for mutual orientation of the polymeric molecules. The multilayers were built up by the usual technique (Blodgett 1935) using very dilute polymer solutions (0·02–0·05 %). As piston oil, castor oil ($F = 17$ dynes/cm.) was employed. At this pressure, however, some of the films are in the

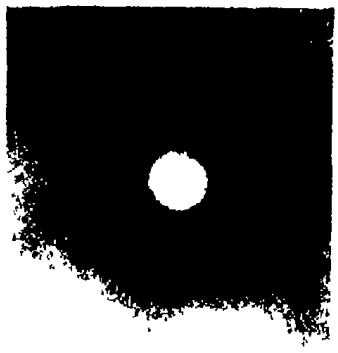


FIGURE 1 A

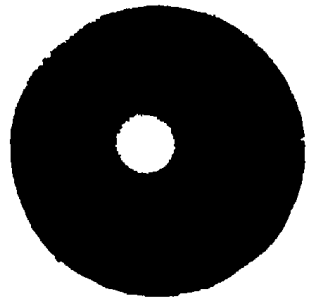


FIGURE 1 B

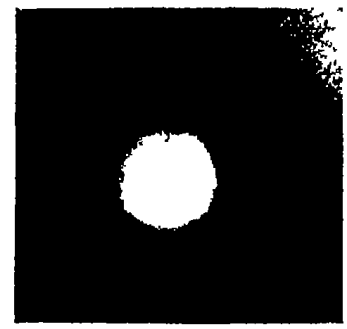


FIGURE 1 C

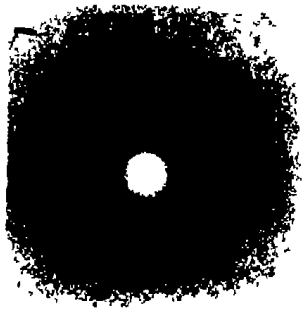


FIGURE 2 A

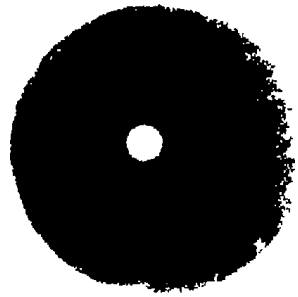


FIGURE 2 B

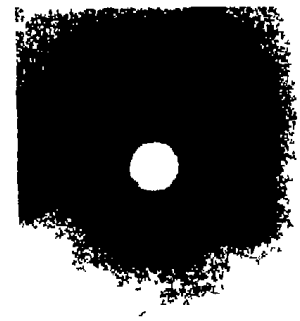


FIGURE 2 C

Photographs of polymer lenses (figure 1), and multilayers (figure 2) (A) polyvinyl acetate (433–482), (B) ethyl-acrylate (416, 548), (C) β ethoxy-ethyl-methacrylate (562, 409).

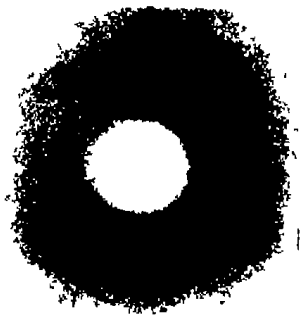


FIGURE 3 A

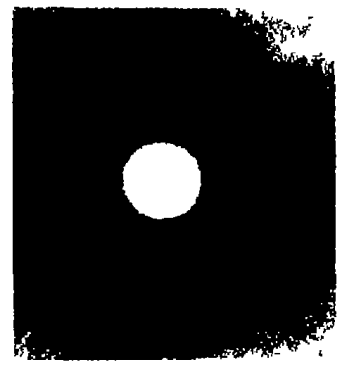


FIGURE 3 B

FIGURE 3 A. Pattern of a cyclo hexyl methacrylate multilayer. The halo d giving the lateral spacing of the side chains becomes dominant and sharper (631).

FIGURE 3 B. A lens of methyl-methacrylate showing that the halo d is absent (382)

(*Facing p.* 168)

gel form (Crisp 1942) and practically all of them have ceased to be perfect monolayers because a fraction of the residues have been forced out of the plane of the surface. Unfortunately, deposition at lower pressures, e.g. with tricresyl phosphate as piston oil, is unsatisfactory owing to a fall in the deposition ratio.

The methyl and cyclohexyl methacrylate monolayers are of a condensed type exhibiting elastic properties at very low pressures and gelate at a pressure of less than 1 dyne/cm. Above 10 dynes/cm. pressure slow collapse takes place. Successful deposition of these monolayers is difficult as the surface is wet on withdrawal of the slide.

The methyl and ethyl acrylates, and polyvinyl acetate give fluid films, comparable to the expanded type given by certain long-chain compounds from which multilayers can be built up by *Y* deposition (Bikerman 1939). These were usually 30–40 monolayers in thickness. Ethyl-, *n*-butyl and *β*-ethoxy-ethyl methacrylates are intermediate in character.

The multilayers were deposited on a nitrocellulose membrane stretched over the specimen holder (Cameron & Coumoulos 1941) The nitrocellulose membrane alone reveals an amorphous pattern whose diffraction features hardly appear after deposition of the multilayer.

Measurement of the haloes

Photographs were taken at several inclinations to the incident electron beam, and in case of the multilayers also in reference to the direction of dipping of the slide.

Measurement of the diameters of the intensity maxima of the haloes was made with the aid of a travelling microscope by the visual method (Pauling & Brockway 1934). Subjective errors, especially in measurement of the first two rings, are liable to be great since the maximum is apparently shifted towards the side of greater contrast, and the intensity of the background varies from the undiffracted central spot to the outside, being steeper as the centre is approached. To eliminate these errors as far as possible photographs were taken with different times of exposure and the background illumination varied to give a suitable contrast. An apparent shift in ring diameter as a function of exposure time was noted and the true diameter determined. Microphotometric recording of the intensities did not prove successful.

The amorphous pattern

The diffraction patterns of all the polymers investigated consist of a series of concentric rings, usually not more than five, which do not alter on heating the specimen or on ageing. The patterns both of the lenses (figure 1, plate 2) and of the multilayers (figure 2, plate 2) were similar in respect to spacings of the corresponding haloes, but these were sharper in the multilayers, one frequently becoming the dominating feature of the photograph, the intensity of the others decreasing.

For different specimens the two inner rings are apparently characteristic, whilst the innermost ring is occasionally missing in the multilayer diffraction pattern. The absence of a large number of rings indicates that the random distribution of scattering centres producing the rings is not in a state comparable to a gas but one similar to a liquid. The photographs are similar to those first obtained by Maxwell (1933) for liquids in which there exists a transient arrangement or a periodicity in distribution of molecules governed by the distances of their nearest approach in their thermal agitation. The absence of the inner haloes in some of the multilayer diffraction patterns is evidence for some degree of orientation—a view supported by the increased sharpness of the halo which becomes the dominating feature in many of the patterns from the multilayers. It is most convenient to compare the relative spacings by utilizing the Bragg formula, although its application to liquids is highly arbitrary. Table 1 shows the values obtained for these spacings. It will be noted that whilst the two outer haloes (a, b) are common to all, the polymers examined in the methacrylate series are distinguished by an extra ring at 2·8A. The dimensions of the two inner rings vary with changes in the nature of the polymer

TABLE 1

				spacings in A (d)		
polymer	(a)	(b)	(c)	lens	multilayer	(e)
methyl methacrylate	1·2	2·2	2·8	—	—	6 6
ethyl methacrylate	1 2	2·2	2 8	4 6	4 7	7 5
n-butyl methacrylate	1·2	2 2	2 8	4·8	4 8	9 0
β-ethoxy-ethyl methacrylate	1·2	2 2	2 8	4 8	4·9	9·5
cyclohexyl methacrylate	—	—	-	5 2	5·3	—
methyl acrylate	1 2	2·2	-	4·1	4·1	6·5
ethyl acrylate	1 2	2·2		4 3	4·3	8 0
polyvinyl acetate (all samples)	1·2	2 2	—	4·0	4 0	7·0

Intensities of halos as $d > c > b > a > e$

The haloes 'a' and 'b'

The two haloes constituting the outer maxima corresponding to Bragg spacings 1·2 and 2·2A appear in all long-chain molecules and are independent of the nature of the polymer. They were first observed by Stewart (1928) in the X-ray diffraction of liquid n-paraffins, and have been reported as present by Wierl (1930) in gaseous hydrocarbons, by Maxwell (1933) and Murison (1934) in oils, by Katzoff (1934) in n-heptane and decane, and by Natta & Rigamonti (1936) in vinyl polymers In a planar zigzag hydrocarbon chain in the gaseous state as noted by Debye (1925) the distance 1·2A corresponds to a Bragg carbon-carbon spacing within one molecule and that of 2·2A to the Bragg distance between a carbon atom and its second neighbour within the molecule (Wierl, 1930), the actual distances are in fact much greater. These haloes are either non-existent or very faint in the multilayers (figure 2, plate 2), suggesting that in these diffraction patterns these haloes are due principally to the diffraction of atoms in the long polymer chain which in the

multilayers lie in planes perpendicular to the electron beam, the more massive side-chains being more or less parallel to the beam do not contribute appreciably to these diffraction maxima.

The halo 'c'

Between the two inner and the two outer haloes the methacrylate polymers alone present an extra halo with a Bragg spacing of 2·8 A which has not yet been reported for any liquid.



FIGURE 4. (*a*) Chain of methacrylate polymers; (*b*) chain of acrylate polymers. The methyl group renders the methacrylate chain less flexible

Since it only appears when a hydrogen atom in the main chain is replaced in each unit by a methyl group, it is suggested that this represents the mean distance between methyl groups along the main chain, considering the methyl groups as separated short side-chains It is interesting to note that a molecular model (figure 4) built to scale yielded a spacing of *ca* 2 8 A for the inter-methyl group separation along the chain This halo is always weakest in comparative intensity, an indication that it may be due to an intramolecular interference

The halo 'd'

Stewart (1930) and his collaborators noted a strong halo at about $d = 4{\cdot}65$ A for the liquid normal paraffins, alcohols and fatty acids, whilst Warren (1933), utilizing a Fourier analysis, obtained a strong peak at $S = \sin\theta/\lambda = 0{\cdot}108$ corresponding to the closest approach of adlineated chains in the liquid normal paraffins Similar spacings have been reported for this halo in a number of polymeric materials, these include

polymer	spacing in A	reference
cumarene	5 3	Beal, Anderson & Long (1932)
indene	5 3	Beal, Anderson & Long (1932)
phenol formaldehyde	4 6	C Ellis (1935)
o-cresol formaldehyde	5 0	C. Ellis (1935)
xylenol formaldehyde	5·5	C. Ellis (1935)
polyvinyl acetate	4·0	Katz (1936)
polyvinyl acetate	3·95	Marvel & Denoon (1938)
polystyrene	4·55	Natta & Rigamonti (1936)

It seems most probable that this spacing corresponds to the lateral spacing between the side-chains pendant on the main polymeric chain. It is this spacing which becomes sharper and more intense, sometimes dominant, in the multilayer photographs (figure 3, plate 2), suggesting that the side-chains lie more or less parallel to the incident electron beam. This halo is absent from the lens photograph of methyl methacrylate (figure 3 A, plate 2). There is, however, a slight trace of the spacing in the multilayer photograph of methyl methacrylate where the side-chains are orientated. In the case of the methyl acrylate (figure 5) the absence of the methyl group from the polymeric chain renders the packing along the chain much less compact and permits of a greater degree of flexibility for the similar short side-chain.

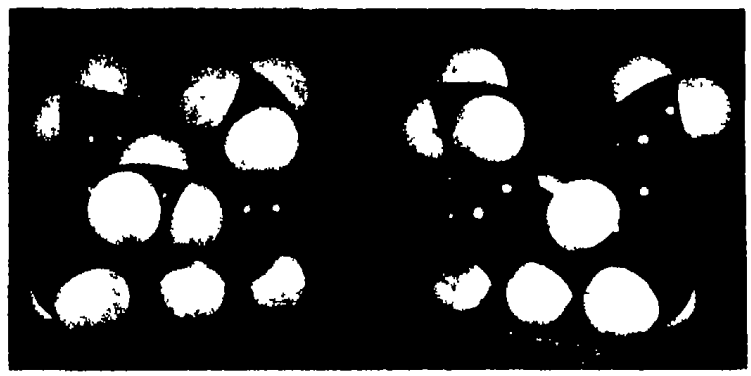
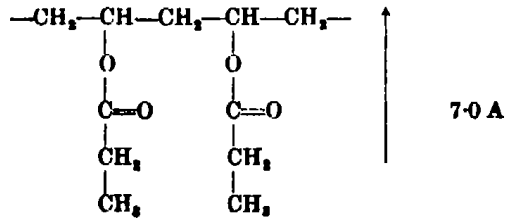


FIGURE 5. Model of methyl methacrylate (*a*), and methyl-acrylate (*b*) polymers. The (*a*) presents a more compact configuration. The side-chains in (*b*) are more flexible.

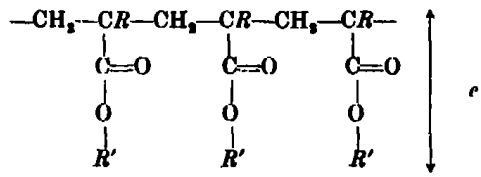
The halo 'e'

This halo is also one which is characteristic of the polymer appearing on polymerization of the monomer. (In general, the dimension corresponds to that given by substances in the liquid state from which the side-chain groups are derived.) It was first observed by Katz (1927) as appearing in some polymerized products and he termed it the 'polymerization ring'. A similar long spacing was first observed by Stewart & Morrow (1927) for straight chain-polar molecules, and they showed that it was a linear function of the number of carbon atoms in the chain Pierce & Macmillan (1938) interpreted similar diffraction maxima in liquid alcohols and acids.

Similar spacings have been observed in the X-ray patterns of several synthetic resins (p. 167). The value of 7·0 A found by Katz (1936) for polyvinyl acetate and confirmed by electron diffraction in these experiments corresponds to what on Staudinger's structure for the polymer may be called the 'height' of the polymer.



The spacing thus corresponds to the distance between the long main chains of the polymeric molecules at their distance of closest mutual approach when separated by one side-chain. We must regard this length as including the carbon atom from which the side-chains hang. Thus in the acrylate and methyl acrylate series we have.



where $R = H$ or CH_3 and $R' = —CH_3$, $—C_2H_5$, $—C_4H_7$, etc. This view is supported by the fact that this long spacing does not usually appear in photographs of the multilayers, and that taking into account the inaccuracy in determining intensity maxima near the undiffracted spot there is good agreement between the lengths of these spacings and the computed lengths of the side-chains (see Table 1).

Discussion

The structure of the polymers

From the above data it is possible to make certain inferences regarding the structure of these polymers

1. The spacings corresponding to the two characteristic 'inner' haloes are due to scattering by the side-chains in positions of closest approach side-to-side, and end-to-end. The end-to-end distance, in this case, is what has been called 'height' of the polymer Thus in these polymers the side-chains are those which are responsible for the observed amorphous pattern and they must be grouped in clusters of molecules, with their chains or rings tending to lie parallel to one another at distances of closest approach (figure 6) in an otherwise random distribution which characterizes the liquid state. These views are not dependent on any assumption that the side-chains which form the cluster belong to the same polymer molecule or to its immediate neighbours.

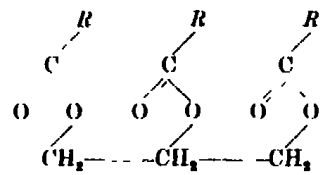
The multilayer pattern reveals a certain degree of orientation of the side-chains, as the 'main inner' halo becomes sharper and the first inner halo sometimes disappears. As has already been mentioned, these polymers do not form regularly built-up multilayers, and so patterns derived from the multilayers give merely indications of orientation of the side-chains However, it may be assumed that these chains lie in planes perpendicular to the main long chains, which in the multilayers are parallel to the base supporting the specimen. The broader halo of the lens patterns indicates that conditions for a regular arrangement of the side-chains are less probable than in the multilayers. Owing to thermal agitation, there is on the average a marked disarray even of the closest neighbours, and at a distance of some molecular diameters there is no reason to assume any fixed orientation.

In a monomolecular film (at high pressures) deposited on water the side-chains orient themselves approximately perpendicular to the water surface. This monolayer behaves as a series of repeated units irrespective of the number of these in one molecule, and each orientates into positions so as to occupy a minimum space compatible with the electrical forces at the interface The pseudocondensed monomolecular films differ from the expanded type in their lack of definite collapse pressure, irreversibility and strong coherence



FIGURE 6. Model of *n*-butyl methacrylate polymer. The side-chains cluster together, parallel to one another

It is clear therefore that a multilayer built from monomolecular films, in which the side-chains are subject to strong field forces on the water surface, will show more exaggerated orientational features, just as in films of the triglycerides the side-chains orientate themselves in the interface adjacent to one another (Alexander 1941)



2. For the structure of the long main chain from which the side-chains hang, we do not possess much experimental evidence apart from the recorded spacings 1·2 and 2·2A of the 'outer' haloes. The structure must be such as to permit of close packing of the side-chains in such an arrangement that their side-to-side and end-to-end spacings approach those calculated from the diffraction patterns. It

appears that the regular zigzag hydrocarbon chain fulfils these conditions. This would give the observed spacings of 1·2 and 2·2A, and bring all the side-chains into favourable positions for their close packing. The side-chains must be attached to every alternate carbon atom of the main chain; this is in accordance with the accepted 1, 3 structure of these polymers. Further, the side-chains alternate on either side of what would be the plane of the zigzag carbon chain (figure 7).

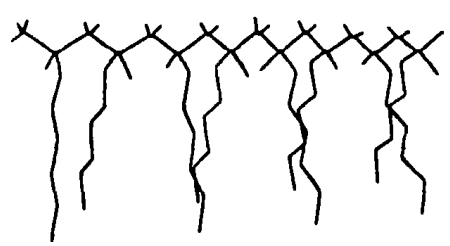


FIGURE 7. Diagrammatic configuration for linear polymers with side-chains

The side-chains clustering together distort the plane of the zigzag carbon main chain. The methyl groups of the methacrylate polymers are also, of course, on alternate sides of the main chain. In the methacrylates, as is shown on models (figure 6) built with Stuart's (1934) model atoms, the distorted chain is a relatively stiff one

This view receives confirmation from Hibbens' (1937), and Monnier, Susz & Briner's (1938) experiments on the Raman spectra of acrylic acid, methyl and ethyl methacrylates at different stages of polymerization The changes in the spectra indicated the disappearance of a $-C\langle^{H}_{H}$ as well as of a —C=C— linkage. Lower frequencies indicated increased chain length. Further, they obtained indications of an increased symmetry of the molecule The force constants of the more characteristic groups approximated more closely to the *cis*- than to the *trans*-modification, and there were also indications of a continuity of structure.

One may also note that Marvel & Denoon (1938) obtained chemical evidence for the preferred 1, 3 structure in vinyl acetate polymers.

3. Adopting as an ideal polyacrylate structure a plane zigzag for the main chain with pendant side-chains in the 1, 3 position perpendicular to and on alternate sides of the main chain, it is possible to calculate the interaction energy between a side-chain and its nearest neighbours (five in number). In this computation the dipole interaction between the pendant ester groups being in *cis* position relative to one another (Alexander & Schulman 1937) was evaluated by means of the equation

$$E = \epsilon_1 \epsilon_2 \sum_{i=j} \frac{1}{d_{ij}},$$

where ϵ_1, ϵ_2 are the charges, and d_{ij} the distances between them. For every (CH_2)

group in the paraffin chain pendant on the ester group the van der Waals interaction was calculated by means of the equation developed by Orr (1942)

$$E = \left\{\frac{-0{\cdot}148}{R^6} + \frac{704}{R^{12}}\right\} 10^{-9} \text{ ergs},$$

where R is the distance between two —CH_2— groups. The results obtained (figure 8) indicate strong repulsive forces between the side-chains on the alternating sides of the main chain until their length is 8 carbon atoms, the contribution of the polar groups being $-2{\cdot}3$ kg. cal./mol.

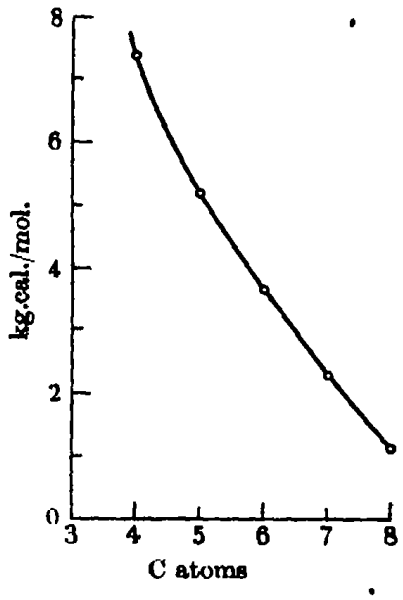


FIGURE 8

Amorphous structure

4. It is impossible to detect any crystalline features in these patterns. Furthermore, the liquid structure of the patterns persists on stretching (Katz 1936). Only in the case of polyvinyl acetate, as reported by Katz (1936), does the inner halo or 'polymerization' ring develop two equatorial maxima in the direction of stretching. These polymers with large side-chains give diffraction patterns characteristic of their side-chains. On stretching, the long main chains are affected by the extending forces more than the comparatively shorter side-chains. This is not the case with the 'linear' polymers, e g. rubber. Their amorphous pattern is due to the long macro-molecules in the liquid state. This amorphous pattern remains, becoming weaker and superimposed, on the crystalline pattern which appears on stretching. The isoprene units of the rubber in the stretched state are extended and oriented in a parallel fashion to produce crystal fibre patterns. It is interesting to compare the behaviour of the different vinyl polymers on stretching. A comparison between polyvinyl acetate and polyvinyl alcohol or halides (figure 9) reveals that the former belongs to the first type with large side-chains and the others are polymers of the linear type.

In the case of the methacrylate polymers the mutual repulsion is greater owing to the presence of the methyl group between the two side-chains. These repulsive forces can be lowered by rotation of the chain. The most probable position is the one which brings the neighbouring side-chains into parallel positions, 4·6 A apart, and the dipoles in positions of least interaction. Furthermore, rotations about the other atoms in the side-chains are equally probable, which will assist close packing parallel to one another.

As in liquids, these polymers are partially ordered arrangements of molecules, or rather of their side-chains, and, of course the shape of these chains plays an important part.

It may be suggested that the amorphous state of these linear polymers with large side-chains is due to the fact that the side-chains adapting themselves to the close-packed form, which is required by the diffraction photographs, distort the main chain The ester groups on the long main chain are jointed at distances of about 2·5Å by the —C—C— link, and it is thus probable that there is more interaction by means of van der Waals forces between the large side-chains than between the main chains which are widely separated from one another (figure 7). Polyvinyl alcohol and halides develop a 'crystalline' pattern on stretching,[5] while polyvinyl acetate retains its amorphous pattern (figure 9)



FIGURE 9 Polyvinyl (*a*) acetate, and (*b*) chloride The side-chains of the polyvinyl acetate are responsible for the amorphous state of this polymer

Structure and elastic properties

5 Methyl methacrylate at ordinary temperatures is slightly extensible and has a high rate of recovery, compared with methyl acrylate which is more extensible but has a lower rate of recovery. In the case of methyl methacrylate except for the 'stiff' long main chain, the side-chains are short and inflexible owing to strong attractions of polar groups $\left(\begin{matrix}\mathrm{O}\\ \|\\ \mathrm{—C—O—}\end{matrix}\right)$ situated opposite the $\mathrm{—CH_3}$ groups of the main chains, and the polymer molecule may be regarded as a stiff cylindrical rod. Similar conclusions have been reached from the analysis of its lens pattern. Methyl acrylate, owing to the absence of the methyl group, has not only a flexible main chain but more flexible side-chains.

This difference in elastic properties coincides generally with the distinction made for the monomolecular films between those of the (1) coherent and (2) fluid or expanded type. Fibres made from monomolecular films (Crisp 1942) show similar elastic properties. These fibres of methyl and cyclohexyl methacrylates are brittle and not extensible, while those of the fluid type are elastic and extensible; the *n*-butyl methacrylate shows a high extensibility, giving a highly viscous almost liquid fibre. The same phenomena were observed in the monomolecular films, while building up the multilayers. On dipping, the coherent films break and the ex-

panded films show higher extensibility and as the side-chains become longer this property is more marked. The cyclohexyl methacrylate whose side-chains are of a more rigid structure presents in general the same structure (figure 10) as the methyl methacrylate. Both polymers have a more compact structure.

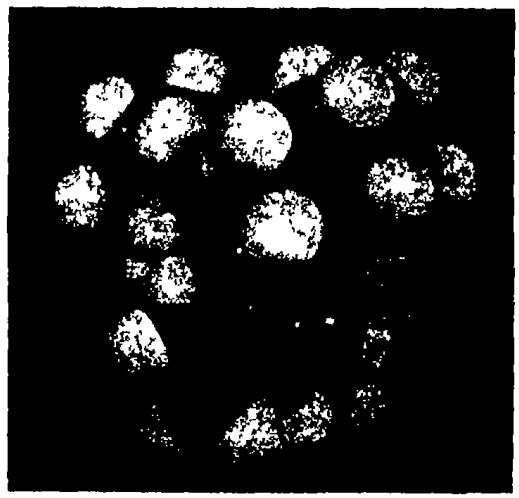


FIGURE 10. Model of cyclohexyl methacrylate showing a compact configuration



FIGURE 11 Coiling of methacrylate chain. Approximately 40 units conclude a whole circle (See also figure 6)

Because of the stiffness of the main chain it is difficult to apply directly to these polymers Kuhn's theory of high elasticity; namely, that the tension exerted by an extended molecule represents the tendency for it to take up the statistically most probable overall length. It may be noted, however, that the methacrylate main chain can coil in a spiral of approximately 65 A diameter (figure 11), the height of the polymer determining the distance apart of the successive loops of the spiral

(figure 6). The extension, then, of this polymer is due chiefly to the rotation of the side-chains, and is approximately proportional to their length. This does not hold for very long side-chains like the octadecyl methacrylate because crystallization takes place at the ends of these side-chains. A smaller proportion of the extension may be due, as well, to the spiral of the main chain. This suggested picture is naturally highly idealized, and merely shows the limited capacity of coiling of the stiff methacrylate chain. The more flexible acrylate chain will have a smaller diameter, and there the main chain will contribute to a higher degree to the high elasticity.

My thanks are due to Professor E. K. Rideal for suggesting the subject, and for his stimulating discussions, also to the British Council for the grant of a scholarship which made this work possible. I wish also to thank the Imperial Chemical Industries, Ltd. for the supply of the polymers.

References

Alexander 1941 *Trans. Faraday Soc.* **37**, 432.
Alexander & Schulman 1937 *Proc Roy. Soc.* A, **161**, 115.
Beal, Anderson & Long 1932 *Industr. Engng Chem.* **24**, 1068.
Bikerman 1939 *Proc. Roy. Soc.* A, **170**, 130
Blodgett 1935 *J. Amer. Chem. Soc.* **57**, 1007.
Cameron & Coumoulos 1941 *Proc Roy Soc.* A, **178**, 415.
Crisp 1942 Private communication.
Debye 1925 *J. Math. Phys* **4**, 133.
Ellis 1935 *Chemistry of natural and synthetic resins.* New York.
Hibbens 1937 *J. Chem. Phys* **5**, 706.
Katz 1927 *Z. phys. Chem.* A, **125**, 321.
Katz 1936 *Trans. Faraday Soc.* **32**, 77.
Katzoff 1934 *J. Chem. Phys.* **2**, 841.
Marvel & Denoon 1938 *J Amer. Chem Soc* **60**, 1045.
Maxwell 1933 *Phys. Rev.* **44**, 73.
Monnier, Susz & Briner 1937 *C R. Soc. phys. Hist. nat. Genève*, **54**, 104.
Murison 1934 *Phil. Mag.* **17**, 201
Natta & Rigamonti 1936 *R.C Accad. Lincei*, (6), **24**, 381.
Orr 1942 Private communication.
Pauling & Brockway 1934 *J Chem. Phys.* **2**, 867.
Pierce & Macmillan 1938 *J. Amer. Chem. Soc* **60**, 779.
Simard & Warren 1936 *J. Amer. Chem. Soc.* **58**, 507.
Stewart 1928 *Phys. Rev.* **31**, 174.
Stewart 1930 *Rev. Mod. Phys.* **2**, 116.
Stewart & Morrow 1927 *Phys. Rev* **30**, 232.
Stuart 1934 *Z. phys Chem.* B, **27**, 350.
Warren 1933 *Phys. Rev.* **44**, 969.
Wierl 1930 *Ann. Phys., Lpz.*, **8**, 521.

The production of penetrating showers

By L. Jánossy and G. D. Rochester
The Physical Laboratories, The University, Manchester

(*Communicated by P. M. S. Blackett, F.R.S.—Received 2 March* 1943)

It is shown that about one-third of the radiation producing penetrating showers is non-ionizing and more penetrating than photons. The total intensity of this non-ionizing radiation (named N-radiation) is found to be about 0·001 % of the full cosmic radiation near sea-level. The N-radiation is possibly the energetic part of the penetrating non-ionizing component of cosmic radiation. It is suggested that this radiation consists of neutrons.

1. Introduction

In a recent investigation Jánossy (1942) showed that the coincidence rate of a counter arrangement which recorded penetrating showers increased appreciably when a block of lead 1·8 cm thick was placed above the top layer of counters. The increase was interpreted as due mainly to the production of penetrating showers in the lead by a shower-producing radiation. In this paper, which gives the results of an investigation into the nature of the shower-producing radiation, it is shown that part of the showers are produced by a non-ionizing radiation (called for brevity N-radiation) more penetrating than photons. The production of penetrating showers by photons was not investigated

2. The experimental arrangement and the results

A. The experimental arrangement is illustrated diagrammatically in figure 1. It consists of thirty-one coincidence counters arranged in three trays B, C, and D and an anticoincidence system of thirty-five counters in parallel. The counters were of the alcohol-argon type, their construction and properties are discussed elsewhere (Rochester & Jánossy 1943).

The counter trays were separated by 15 cm. of lead and were surrounded by a lead absorber at least 50 cm. thick.

An absorber T was placed close above the counters B and was shielded from all sides except the bottom by the anticoincidence counters A. An absorber Σ was placed close above the anticoincidence counters A.

Each anticoincidence counter had an efficiency greater than 99 %.

Sevenfold coincidences $B_1B_2B_3C_1C_2D_1D_2$ (which will subsequently be abbreviated $B_{123}C_{12}D_{12}$) were recorded. A sevenfold coincidence indicated that at least three counters from tray B and two from each of the trays C and D were discharged.

Simultaneously with the coincidences, anticoincidences $B_{123}C_{12}D_{12}-A$, i.e. sevenfold coincidences not accompanied by the discharge of any of the counters A, were recorded.

Most of the coincidences and almost all of the anticoincidences were caused by penetrating showers (see § 3B).

B. The purpose of the experiment, as will be explained more fully in § 3A, was to show that the rate of anticoincidences for a given value of T decreased with increasing thickness of the absorber Σ. To this end readings were taken of the rates with $\Sigma = 0$ cm., 5 cm. and 35 cm. of lead.

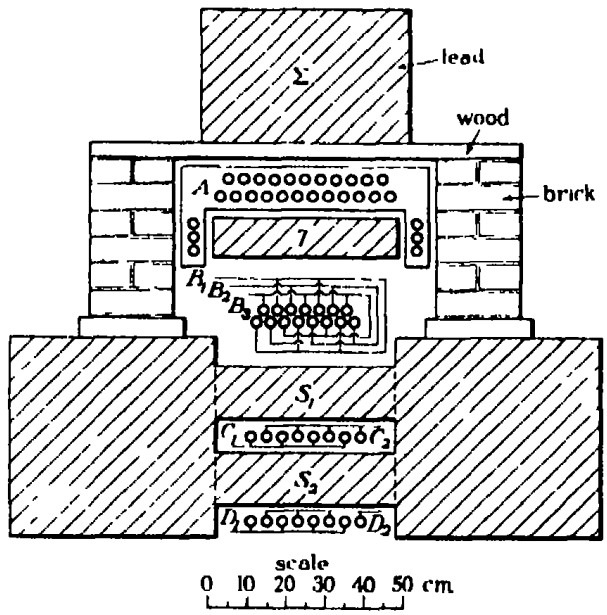


FIGURE 1. The experimental outlay

The results of five months continuous recording are given in table 1 and figure 2. The anticoincidence rates are corrected for random coincidences. This correction is positive because the accidental overlappings of discharges of the counters A with genuine sevenfold coincidences causes a loss of counts

TABLE 1

absorbers			coincidences		anticoincidences	
T	Σ	time	total	rate	total	rate*
cm. Pb	cm. Pb	hr.	no.	per hr.	no.	per hr.
0	0	306.15	162	0·529 ± 0·042	5	0·023 ± 0·010
10	0	380.30	311	0·818 ± 0·046	24	0·086 ± 0·018
10	5	1289.56	1100	0·853 ± 0·026	59	0·063 ± 0·008
10	35	1344.20	924	0·688 ± 0·023	33	0·033 ± 0·006

* Corrected for random coincidences.

The anticoincidence rates given in the last column are very small, the lowest rate observed being only one anticoincidence in 2 days. Nevertheless, there seems to be no reason to doubt the accuracy of the results, firstly because the rates given represent the averages of large numbers of consistent single observations, and

secondly, because the effects though small in absolute magnitude are relatively large.

It will be observed that the longest series of readings was carried out to determine the difference in the rates with $\Sigma = 5$ cm. Pb and $\Sigma = 35$ cm. Pb with $T = 10$ cm. Pb To eliminate fluctuations other than statistical, about forty individual readings were taken for these thicknesses of Σ. The difference between the anticoincidence rates is seen to be $0{\cdot}030 \pm 0{\cdot}010$ counts per hour, and since it is equal to three times the standard error it is likely to be real

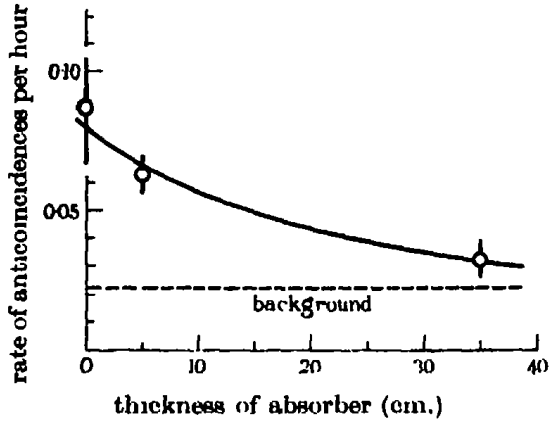


FIGURE 2. Anticoincidence rate for different thicknesses of Σ

Less time was spent in establishing the difference between the rates for $\Sigma = 0$ and $\Sigma = 35$ cm. Pb, but the difference obtained, $0{\cdot}053 \pm 0{\cdot}019$ counts per hour, is almost three times the standard error and therefore supports the conclusion that the anticoincidence rate decreases with increasing thickness of the absorber Σ.

3. THE DETECTION OF THE N-RADIATION

A The method of detection of the N-radiation was similar in principle to that used by Jánossy & Rossi (1940) for the investigation of photons and by Rossi & Regener (1940) and Jánossy & Rochester (1941, 1943) in the investigation of penetrating non-ionizing radiation.

N-rays falling on the absorber T produce penetrating showers and thus give rise to anticoincidences $B_{123}C_{12}D_{12} - A$. Anticoincidences can, however, also be caused by processes not involving N-rays, for example, by side-showers which miss the counters A

The following argument shows that some of the anticoincidences are due to N-rays. The N-radiation falling on T must pass through Σ and is absorbed there. Thus with increasing thickness of Σ the rate of anticoincidences caused by N-rays is expected to decrease. Anticoincidences due to side-showers, however, would hardly be affected by the thickness of Σ. Spurious anticoincidences due to causes other than side-showers will be shown in § 3C to be almost negligible, and hence

their change with Σ must also be negligible. Thus the change in the anticoincidence rate with Σ can be regarded as a measure of the frequency of the N-radiation. As the observations (table 1) show a marked decrease in the anticoincidence rate with increasing Σ, we conclude that the observations show that the N-radiation gives rise to penetrating showers in the absorber T.

B. *Interpretation of coincidences*

In this section we discuss processes other than penetrating showers which are capable of producing sevenfold coincidences

(1) *Cascade showers* Jánossy (1942, p 364) has shown that cascade showers of sufficient energy to produce coincidences in an arrangement similar to ours are too rare to be of importance These arguments can be applied to our present arrangement as well.

(2) *Triple knock-ons* A single penetrating particle might give rise to three independent secondaries, one near each of the counter trays B, C and D, and thus give rise to a sevenfold coincidence by a triple knock-on process The rates of triple knock-ons were measured as described by Jánossy (1942, p. 366). The relevant figures are collected in table 2

TABLE 2

description	symbol	rate per hr. $T=0$	rate per hr. $T=10$ cm. Pb
threefolds BCD	R_3	6438	—
fourfolds $BC_{12}D$	R_4'	211	—
fourfolds BCD_{12}	R_4''	100	—
fivefolds $B_{123}CD$	R_5	140	192

Rate of triple knock-ons $= [R_4' \; R_4'' \; R_5]/R_3^2 = \begin{cases} 0\cdot07 \text{ c p hr} & \text{for } T=0 \text{ cm. Pb} \\ 0\cdot10 \text{ c p hr.} & \text{for } T=10 \text{ cm. Pb} \end{cases}$

In the above table B means counters B_1, B_2 and B_3 in parallel, etc., BC means counters B in coincidence with counters C, etc.

We note that only 12–14 % of the sevenfold coincidences were due to triple knock-ons

(3) *Casual coincidences.* The largest contributions to the casual coincidences were found to be due to the following overlaps

(*a*) A coincidence between a single knock-on in the top tray (e g. $B_{123}C_1D_1$) and a double coincidence (e g. C_2D_2)

(*b*) A coincidence between a double knock-on (e g. $B_{123}C_1D_{12}$) and a discharge in a single counter (e g. C_2)

We note that both processes (*a*) and (*b*) involve at least one penetrating particle traversing all three counter trays. The rates of casual coincidences not involving such a penetrating particle are completely negligible.

The resolving time of the arrangement was measured and found to be about 7×10^{-5} sec. The casual rates were 0·08 and 0·09 counts per hour for $T = 0$ and $T = 10$ cm. Pb respectively.

The rate of penetrating showers produced in T is thus 0·25 count per hour (see table 3).

TABLE 3

	rates per hr.	
	$T = 0$	$T = 10$ cm. Pb
rate of sevenfold coincidences	0 53	0·82
background, i e. triple knock-ons plus casual coincidences	0·15	0·19
rates due to pentrating showers	0 38	0·63
rate of penetrating showers produced in T	0·25	

C. *Interpretation of Anticoincidences*

In this section we consider three questions: (1) How many anticoincidences are recorded which are not connected with penetrating showers? (2) How many anticoincidences are spurious, i e not due to N-radiation? (3) How often is a penetrating shower produced by N-radiation and recorded as a coincidence but not as an anticoincidence?

(1) Though some 25 % of the sevenfold coincidences are caused by processes other than penetrating showers, practically all anticoincidences are caused by penetrating showers. It was shown above that all important processes causing sevenfold coincidences are due either to a penetrating shower or to at least one penetrating particle traversing the three counter trays B, C and D This penetrating particle is likely to discharge the anticoincidence system and thus prevent an anticoincidence.

Anticoincidences not caused by penetrating showers thus occur either in conjunction with a sevenfold coincidence of one of the rare types which do not contain any penetrating particles or in conjunction with a sevenfold coincidence which does contain a penetrating particle which happens not to discharge the anticoincidence due to the inefficiency of the counters Both processes are very rare and it can be shown that together they give rise to a rate less than 1/100 of the total anticoincidence rate.

Another possibility is that a single penetrating particle produced by the N-radiation in T might give rise to a triple knock-on. This process is very rare for the following reason. The probability of a penetrating particle giving rise to a triple knock-on is, according to table 2, of the order of 1/60,000. Thus to account for the whole of the anticoincidence rate of 0·03 count per hour in terms of triple knock-ons a flux of at least $0\ 03 \times 60{,}000 = 1800$ non-ionizing particles per hour is required. This rate is comparable with the total cosmic ray intensity. It has been shown by Jánossy & Rochester (1943) that the total flux of penetrating non-ionizing rays is only of the order of 0·1 % of the cosmic-ray intensity, hence the

rate of triple knock-ons started by non-ionizing rays cannot exceed 10^{-4} per hour and is therefore negligible.

In conclusion, we note that practically all anticoincidences are connected with penetrating showers and hence it is legitimate to assume that the N-radiation absorbed in T gives rise mainly to penetrating showers.

(2) *Spurious anticoincidences*, i.e. anticoincidences not caused by a non-ionizing radiation, can be due to the following processes:

(i) A penetrating shower coming from the side and missing the anticoincidence counters. The rate of such anticoincidences is not likely to be affected by the absorbers T or Σ and thus is probably of the order of the rate of anticoincidences observed with $T = 0$ and $\Sigma = 0$ which is the smallest anticoincidence rate observed.

(ii) Spurious anticoincidences are recorded if a penetrating shower traversing A fails to discharge the anticoincidence system because of the inefficiency of the counters. The probability that a penetrating shower coming from above the apparatus is missed in this way is negligibly small, as most penetrating showers will traverse more than one of the counters A simultaneously. The rate of sevenfold coincidences due to processes containing only one ionizing particle falling on T is composed of two parts (1) the background coincidences (table 3), (2) coincidences due to showers produced in T by ionizing particles. This rate does not exceed $0{\cdot}19 + 0{\cdot}25 = 0{\cdot}44$ count per hour for $T = 10$ cm Pb according to table 3. Assuming that $\frac{1}{2}$ % of these coincidences are recorded as anticoincidences due to inefficiency of A, we obtain 0·002 spurious anticoincidences per hour. This estimate is probably on the high side Thus we conclude that the rate of spurious anticoincidences due to inefficiency is negligible.

(iii) A shower travelling upwards if stopped in T would give rise to a spurious anticoincidence. Such a process must be regarded, however, as extremely unlikely Nevertheless, should such showers occur their rate would not be affected by Σ and thus they would not affect our conclusions.

As we are not aware of any further processes giving rise to spurious anticoincidences, we conclude that spurious anticoincidences other than those due to side-showers are negligible.

(3) A penetrating shower produced in T by the N-radiation is only recorded as an anticoincidence if (i) it does not contain particles moving upwards which discharge the counters A, (ii) if the N-ray is not accompanied by an ionizing secondary. As both processes may occur, the rate of anticoincidence provides only a lower limit to the actual rate of N-rays.

The upper limit of the rate of N-rays is given by the number of penetrating showers produced in T.

4. The interpretation of the results

A. *The range of the* N-*radiation*

Since the rate of anticoincidences decreased considerably when Σ was changed from 5 to 35 cm. Pb, the N-radiation must have a range exceeding 5 cm. Pb. The

radiation is therefore more penetrating than photons. It cannot be concluded from this that photons do not produce penetrating showers, for some of the anticoincidences observed with $\Sigma = 0$ might well be due to photons. It may be noted, however, that even if photon-produced penetrating showers exist, an anticoincidence arrangement would be a very inefficient recorder. The reason for this is that only energetic photons would be capable of producing penetrating showers, and these would be accompanied by large numbers of ionizing particles which would prevent an anticoincidence from being recorded.

Thus we conclude that the penetrating showers observed are produced by non-ionizing radiation more penetrating than photons, but that the production of penetrating showers, even in large numbers, by photons cannot be excluded.

B. *The fraction of penetrating showers produced by* N-*radiation*

An anticoincidence will be produced by an N-ray only if the following conditions are fulfilled: (i) the N-ray must traverse Σ and must not produce in Σ a penetrating shower or ionizing secondaries, (ii) the N-ray must produce a penetrating shower while traversing T, and (iii) the penetrating shower produced by the N-ray must give rise to a coincidence $B_{123}C_{12}D_{12}$

Assuming that the absorption of the N-radiation in lead takes place mainly by the production of penetrating showers, the probability of an N-ray being recorded can be written as

$$P(\Sigma, T) = e^{-\mu\Sigma}(1 - e^{-\mu T})\, P_0, \tag{1}$$

where μ is the absorption coefficient of the N-radiation and P_0 is the probability that a penetrating shower produced in T is recorded as a sevenfold coincidence. In the absence of precise information about the value of P_0 we assume that $P_0 = 1$.

Values of $P(\Sigma, T)$ for $\Sigma = 5$ cm. Pb and $T = 10$ cm. Pb are given in table 4 We note that P is of the order of 30 % for a large range of μ values, and therefore whatever the actual value of μ only about one-third of the N-radiation can be recorded with our arrangement.

TABLE 4

$1/\mu$ cm. Pb	P %	$1/\mu$ cm. Pb	P %
5	31·8	15	34·8
10	38·5	20	30·8

It is known from previous work (Jánossy 1942, p. 371) that the bulk of the radiation producing penetrating showers is of short range. Thus it can be assumed that most of the radiation other than N-radiation falling on T is absorbed, producing penetrating showers. Writing N_0 for the intensity of the N-radiation falling on T, I_0 for the intensity of all other radiations which produce penetrating showers, and R for the rate of all penetrating showers produced in T, we have

$$R = I_0 + N_0(1 - e^{-\mu T}). \tag{2}$$

The rate of anticoincidences can be written

$$A = N_0 e^{-\mu\Sigma}(1 - e^{-\mu T}) + B, \tag{3}$$

where B is the background rate. Examination of tables 1 and 3 shows that

$$\left.\begin{aligned} R = 0{\cdot}25 \text{ count per hour}, \quad A(\Sigma = 5\,\text{cm.Pb}) = 0{\cdot}063 \text{ count per hour} \\ \text{and} \qquad A(\Sigma = 35\,\text{cm.Pb}) = 0{\cdot}030 \text{ count per hour.} \end{aligned}\right\} \tag{4}$$

Eliminating B by taking the difference between the values of $\Sigma = 5$ and $\Sigma = 35$ cm. Pb, we can obtain N_0 as function of μ and then from equation (2) I_0/N_0 as a function of μ. The results are given in table 5, from which it is seen that I_0/N_0 is approximately 2.

TABLE 5

$1/\mu$ cm.	N_0 c per hr.	I_0/N_0	$1/\mu$ cm.	N_0 c. per hr.	I_0/N_0
5	0·104	1·63	15	0·109	1·78
10	0 091	2·12	20	0·139	1·40

C. *The absolute intensity of the* N*-radiation*

The estimation of the absolute intensity of the N-radiation is necessarily more uncertain than the estimate of the relative intensities for the following reasons. (1) it is difficult to estimate the solid angle of the sevenfold set of counters effective in the collection of N-radiation, (2) it is not known if the N-radiation is accompanied by ionizing radiation, if there is such N-radiation, it will not be recorded by our arrangement; (3) the assumption $P_0 = 1$ may be incorrect

It is seen from table 5 that N_0 is approximately 0·1 count per hour, and hence the ratio of this to the total cosmic-ray flux through the trays B and C is

$$100 \times 0{\cdot}1/12{,}000 = 0{\cdot}001\ \%.$$

This value gives the order of magnitude of the intensity of the N-radiation in terms of the total cosmic-ray intensity at sea-level

D. *The nature of the* N*-radiation*

(1) Since the intensity of the N-radiation is much less than the intensity of the penetrating non-ionizing radiation given in a former publication (Jánossy & Rochester 1941, 1943), it is an open question whether the two radiations are the same. We would point out, however, that the experiments are compatible with the assumption that the two radiations are the same. The radiation investigated formerly was found to have a mean range of 10 cm. Pb. In the present investigation we find that the N-radiation is considerably absorbed between 5 and 35 cm. Pb, a result compatible with the assumption that the range of the N-radiation is also of the order of 10 cm. Pb.

A difference in the two observed intensities is certainly to be expected if the N-radiation and the non-ionizing radiation observed before are of the same nature,

for the present arrangement responds only to the part of the non-ionizing radiation which has enough energy to produce penetrating showers, whereas in the former experiment non-ionizing particles capable of producing secondaries down to an energy of 2×10^8 eV were observed

(2) It was suggested by Jánossy (1942, pp. 373, 375) that penetrating showers are possibly produced by protons As mesons are produced by non-ionizing radiations at high altitudes (Schein and co-workers 1939, 1940) it was further suggested that penetrating showers might also be produced by neutrons.

This picture fits in with our present observations which show that the penetrating showers are produced by both ionizing and non-ionizing radiations of comparable intensities.

We note that the N-radiation having a range of the order of 10 cm. Pb is less absorbed than the ionizing shower-producing radiation which has a range of the order of a few cm. of lead only. Thus if the two radiations are to be identified with neutrons and protons, one must assume a somewhat different interaction for fast collisions for the two radiations

(3) It should be noted that Johnson (1938, 1939) put forward arguments in support of the view that mesons are produced by protons This view was elaborated by Blackett (1941). According to theoretical considerations by Hamilton, Heitler & Peng* (1943) penetrating showers are expected to be produced almost equally by protons and neutrons. The theoretical cross-section is of the right order of magnitude to account for the rate of production of penetrating showers.

References

Blackett, P. M. S. 1941 *Proc. Phys. Soc* **53**, 203
Hamilton, Heitler & Peng 1943 (in the Press)
Jánossy, L. 1942 *Proc. Roy. Soc.* A, **179**, 361.
Jánossy, L & Rochester, G D 1941 *Nature, Lond*, **148**, 531
Jánossy, L & Rochester, G. D. 1943 *Proc Roy. Soc.* A (in the Press)
Jánossy, L. & Rossi, B. 1940 *Proc Roy. Soc* A, **175**, 88
Johnson, T H 1938 *Phys. Rev.* **53**, 499.
Johnson, T. H 1939 *Rev. Mod. Phys.* **11**, 208.
Rochester, G. D & Jánossy, L 1943 *Phys Rev.* **63**, 52
Rossi, B. & Regener, V H. 1940 *Phys. Rev.* **58**, 837
Schein, M. & Wilson, V. C. 1939 *Rev. Mod. Phys* **11**, 292.
Schein, M , Wollan, E. O. & Groetzingen 1940 *Phys. Rev.* **58**, 1027.

* We are very much indebted to Prof. W. Heitler for a summary of his unpublished results.

The scattering and polarization of fast electrons by heavy elements

By C. B. O. Mohr, Ph.D., *Lecturer in Physics, University of Cape Town*

(*Communicated by H S W. Massey, F R S —Received 2 March* 1943)

Using the methods of a previous joint paper, the calculation of the scattering and polarization of electrons with energies between 5 4 and 1060 keV is extended to the whole angular range for the atomic field of gold It is found that even at the highest energies the angular distribution of the scattering does not exhibit a monotonic fall with increasing angle, and that the asymmetry in double scattering reaches a maximum value which is quite large at angles greater than 90° The variation of the total elastic cross-section with energy is also investigated.

A simple closed formula for the scattering is obtained by transforming Dirac's equations into a single differential equation of the same form as Schrödinger's equation but with a modified 'scattering potential', and then using this potential in Born's approximation The resulting formula gives angular distributions to which the exact values of the scattering appear to approach more and more closely as the energy of the electrons increases

In connexion with the discrepancies between the results of various observers for the scattering and the polarization at 90°, it is shown that if a long-range repulsive field is added to the field of the atom so as appreciably to reduce the scattering, the asymmetry in double scattering is reduced by a larger factor.

While the theory of the interaction of electrons with matter has been remarkably successful in treating the fast electrons met with in cosmic rays, certain observers have claimed that serious discrepancies occur for electrons in the energy range 100–2000 keV. The experiments of Dymond (1934) and of Richter (1937) failed to reveal anything like as large a polarization as predicted by the calculations of Mott (1932), while the experiments of Klarmann & Bothe (1936) and of Champion & Barber (1938) indicated that the intensity of large-angle scattering of electrons by heavy electrons was only about one-sixth of that given by theory (Mott, 1929). On the other hand, the scattering experiments of Gupta (1939) do not give results in disagreement with theory for electrons within the energy range used by Klarmann & Bothe, while the recent experiments of Shull (1942) on the polarization of 400 kV electrons by gold show that with certain experimental conditions an asymmetry in double scattering of the order of magnitude predicted by Mott is obtained.

In view of the discrepancy between the results of Dymond's experiments and of Mott's calculations, Massey and the writer showed in a previous paper, hereafter referred to as Paper A (1941), that the calculated polarization at 90° for the atomic field of gold is nearly the same as that calculated by Mott for the Coulomb field of the gold nucleus The same result was arrived at almost simultaneously by Bartlett & Watson (1940) for the Coulomb field of the mercury nucleus, and by Bartlett & Welton (1941) for 100 and 230 kV electrons and the atomic field of mercury.

The present paper gives the results of an investigation of the scattering and polarization of electrons over a wide range of energies and angles, for the atomic field

of gold. It is found that diffraction minima persist in the angular distribution of the scattering even at the highest energies, though otherwise the curves seem to approach more and more closely to values given by a simple closed formula for the scattering derived from Dirac's equations using Born's approximation. The polarization is a maximum, not at 90°, but at a larger angle, and the magnitude of the asymmetry in double scattering is then quite large. Finally, the relative effects of an alteration in the field on the scattering and on the polarization are considered.

General theory

Consider the scattering through an angle θ of an unpolarized beam of electrons of rest mass m, velocity v, energy E, and wave number k, so that

$$k^2 = \frac{4\pi^2}{h^2}\left(\frac{E^2}{c^2} - m^2c^2\right) = \frac{4\pi^2 m^2 v^2}{h^2}\left(1 - \frac{v^2}{c^2}\right)^{-1} \tag{1}$$

As shown in Paper A, the effective cross-section for scattering per unit solid angle in the direction θ is given by

$$I(\theta) = |f(\theta)|^2 + |g(\theta)|^2, \tag{2}$$

where

$$\left.\begin{aligned} f(\theta) &= \frac{i}{2k}\Sigma\{(l+1)(1-e^{2i\eta_l}) + l(1-e^{2i\eta_{-l-1}})\}P_l(\cos\theta), \\ g(\theta) &= \frac{i}{2k}\Sigma\{e^{2i\eta_l} - e^{2i\eta_{-l-1}}\}P_l^1(\cos\theta). \end{aligned}\right\} \tag{3}$$

The quantity η_l is a phase constant such that $\sin(kr - \frac{1}{2}l\pi + \eta_l)$ is the asymptotic form of that solution of the differential equation

$$\frac{d^2G_l}{dr^2} + \left\{\alpha\beta - \frac{l(l+1)}{r^2} + \frac{l+1}{r}\frac{\alpha'}{\alpha} - \frac{3}{4}\frac{\alpha'^2}{\alpha^2} + \frac{1}{2}\frac{\alpha''}{\alpha}\right\}G_l = 0, \tag{4}$$

which vanishes at the origin, the quantities α and β being given by

$$\alpha = \frac{2\pi}{hc}(E - V + mc^2), \quad \beta = \frac{2\pi}{hc}(E - V - mc^2).$$

The quantity η_{-l-1} is similarly defined, the corresponding differential equation for G_{-l-1} differing from that for G_l only in having a term $-l\alpha'/r\alpha$ in place of the term $(l+1)\alpha'/r\alpha$ in (4). It turns out that the difference between the values of η_l and η_{-l-1} is negligible except for the smallest values of l, and it is this difference which is responsible for the existence of a polarization on scattering.

Consider now the effect of polarization of the electrons in a double scattering process. If the original beam, incident in the direction L_1T_1, is first scattered through an angle θ in the direction T_1T_2, and then scattered a second time through the angle θ in the direction T_2L_2, and if ϕ is the angle between the planes $L_1T_1T_2$ and $T_1T_2L_2$, the intensity of the doubly scattered beam is proportional to $1 + \delta\cos\phi$, where

$$\delta = |f(\theta)g^*(\theta) - f^*(\theta)g(\theta)|^2/(|f(\theta)|^2 + |g(\theta)|^2)^2. \tag{5}$$

The quantity 2δ is called the asymmetry in double scattering.

Method of calculation

It is first necessary to calculate the phases η_l, η_{-l-1} by accurate numerical integration of the differential equation (4) for small values of l, the higher order phases being calculable with sufficient accuracy by the use of Jeffreys's approximation. Details of the method are given in Paper A, where the behaviour of the phases for the atomic field of gold is illustrated for values of k up to 150 atomic units (246 kV electrons).† The calculation of the phases has now been extended up to $k = 400$ atomic units (1060 kV electrons), and a few values of $\delta_l = \frac{1}{2}(\eta_l + \eta_{-l-1})$ are given in table 1 to show the behaviour of the phases at these higher energies.

Table 1. Values of the phases for scattering by a gold atom

ka_0	kV	δ_0	δ_1	δ_2	δ_3	δ_5	δ_{10}	δ_{20}	δ_{40}	δ_{100}	δ_{200}
100	121	3·39	3·12	2·49	2 09	1·68	1·02	0·58	0·26	0·01	0·00
200	392	3·47	2 83	2·26	1 92	1·61	1 15	0·76	0·42	0·12	0·03
400	1060	3·35	2 81	2·18	1 96	1·76	1·47	1·05	0·67	0·31	0·12

The phases given by Bartlett & Welton (1941) for 100 and 230 kV electrons in mercury check up very satisfactorily with those obtained in the present work, though the former are slightly greater, as would be expected for the slightly larger atom.

The small order phases reach a broad minimum at about $ka_0 = 100$ and then increase again For still larger values of k the increase becomes logarithmic.

Having inserted the values of the phases in equations (3) for f and g, the intensity of single scattering and the polarization by double scattering are calculated from (2) and (5) respectively. A difficulty arises from the fact that at the higher energies the convergence of the series for f is extremely slow. This is not so serious at 90° (Paper A), where the alternate non-vanishing terms in the series (3) are of opposite sign, so that a fairly reliable estimate can be made of the contribution from the higher order terms without much difficulty In the present calculations it was found necessary to calculate the values of P_l and P_l^1 for a number of angles up to $l = 40$ using the recurrence formulae. Furthermore, the procedure, adopted also by Bartlett & Welton, was followed of fitting the coefficients of the higher P_l and P_l^1 with empirical formulae involving the sum of terms of the form $ae^{-\lambda l}$. The relation

$$\sum_{l=0}^{\infty}(2l+1)\,e^{-\lambda l}P_l(\cos\theta) = (1-e^{-2\lambda})/(1-2e^{-\lambda}\cos\theta+e^{-2\lambda})^{\frac{1}{2}}$$

was then used to sum the series, allowance being made for the deviations of the coefficients of the lower order terms from the value given by the empirical formula. The series for g converges much more rapidly, because $\eta_l - \eta_{-l-1}$ decreases rapidly with increasing values of l.

† The values of the δ_l given in figure 1 of Paper A are slightly too large at $ka_0 = 150$ owing to a small systematic error in the calculations, but are correct for ka_0 less than 100. The corrected values of the percentage asymmetry are not, however, seriously different from those given in figure 2 of Paper A.

be said of the Mott values at 90° for the higher energies. For velocities approaching that of light, Mott's value of R tends to 3·4, formula (15) gives 2·7, while the values given by Bartlett & Watson tend to 1·9.

Even above 1060 keV, marked deviations from a monotonically falling angular distribution occur for angles greater than 90°. This is due to the fact that the lowest

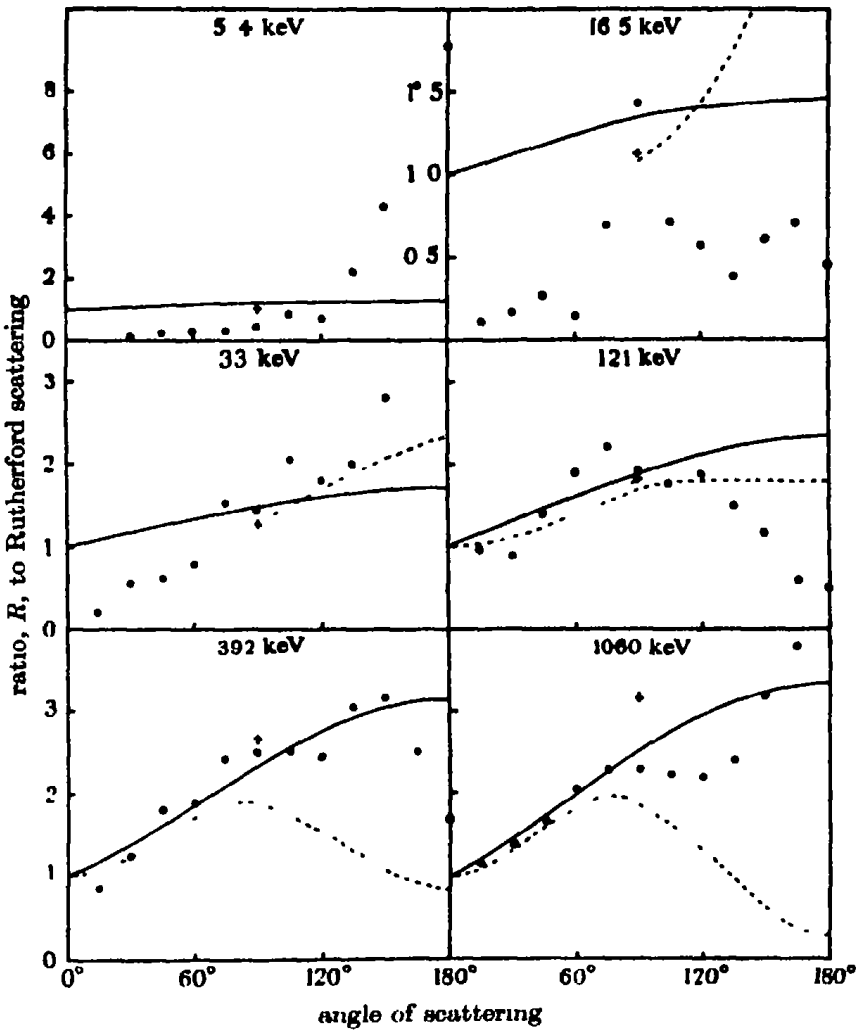


FIGURE 1. Ratio, R, of the scattering of fast electrons by heavy atoms to the Rutherford scattering. Circles, calculated for a screened gold nucleus. Full lines, values given by the approximate closed formula (15). Dotted lines, calculated by Bartlett & Watson for an unscreened mercury nucleus. Crosses, calculated by Mott for an unscreened gold nucleus.

order phases are still widely separated, so that the values of $1-\cos 2\delta_l$ and $\sin 2\delta_l$ do not vary regularly with l for the first few values. Moreover, it appears that this behaviour will persist up to energies of several million electron volts.

Experimental evidence concerning the angular distribution of the scattering of β-particles by heavy elements is so far very sparse for the larger angles, and is even

conflicting in regard to the order of magnitude of the scattering. Klarmann & Bothe (1936) find for 500–2500 kV electrons in krypton and xenon only one-fifth of the scattering given by Mott's formula for light elements (Mott 1929), and similarly Champion & Barber (1938) find for 600–1100 kV electrons in mercury only one-seventh of the theoretical scattering On the other hand, Gupta (1939) finds agreement with theory for the scattering of 2 million volt electrons in xenon. Further

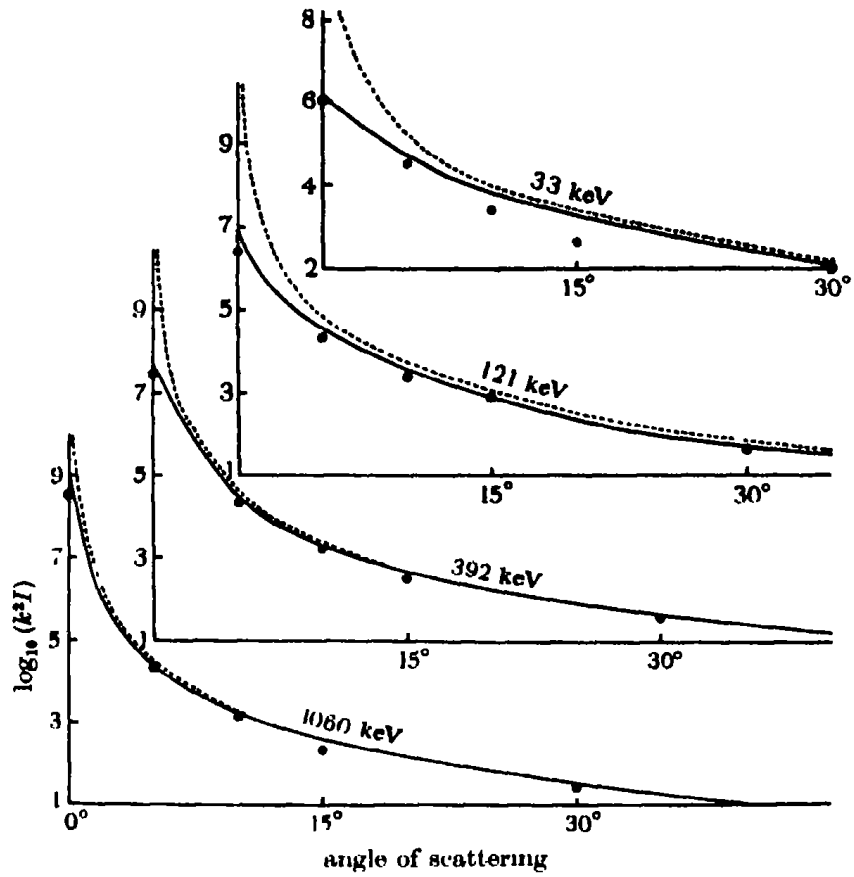


FIGURE 2. Values of $k^2 \times$ (scattered intensity I), plotted on a logarithmic scale, for small angles of scattering. Circles, calculated for a screened gold nucleus Full lines, calculated from formulae (7) and (16) for a screened gold nucleus. Dotted lines, calculated from the corresponding formula (13) for an unscreened gold nucleus.

experiments are clearly desirable not only to clear up the discrepancy in the order of magnitude of the scattering, but also systematically to determine the form of the angular distribution of the scattering at large angles. It may be noted in passing that since Mott's formula for the large-scale scattering by light elements, with which the experimental values have usually been compared, gives much smaller values than the exact values at high energies, the discrepancy between theory and the experiments of Klarmann & Bothe and of Champion & Barber is even greater than at first appears.

With regard to the scattering at small angles, it will be seen on referring again to figure 1 that the exact value of R is obviously tending to smaller values than 1 as the angle decreases. This is to be expected, for the scattering by a Coulomb field tends to infinity as the angle tends to zero, whereas the true value is of course finite for zero angle. At small angles, therefore, a much better approximation to the scattering would be given by taking the atomic field and substituting the value

$$V = Z_p e^2/\{r(1-v^2/c^2)^{\frac{1}{2}}\} + Z_p^2 e^4/2mc^2 r^2 \tag{16}$$

in formula (7), the integration being carried out numerically.

In figure 2 are shown angular distributions so calculated from the formula (7), together with corresponding ones obtained from the formula (13) for the Coulomb field, and the exact values obtained by the use of the formulae (2) and (3). While the exact values deviate from the formula (7) by a factor of 4 at some angles at the lower energies, they never deviate by a factor of more than 2 at the higher energies, and this is a close enough fit for all practical purposes in view of the extremely rapid variation of the intensity with angle. The values given by the simpler formula (13) are seen to hold down to smaller and smaller angles as the electron energy increases, the deviation for 1060 kV electrons only becoming marked for angles less than 1°.

Total elastic cross-section

Since the main contribution to the total cross-section comes from small angles where g is negligible in comparison with f, the total cross-section will be given by

$$\begin{aligned} Q &= 2\pi \int_0^\pi f^2 \sin\theta \, d\theta \\ &= 4\pi k^{-2} \sum_l \{(l+1)\sin^2\eta_l + l\sin^2\eta_{-l-1}\}. \end{aligned} \tag{17}$$

Values were calculated from this formula and the results are given in table 2.

Table 2. Values of the total cross-section for elastic collisions in units of πa_0^2

Energy in keV	5·4	16·5	33	121	392	1060	∞
Cross-section, formula (17)	1·42	0 86	0·40	0 21	0·12	0·11	—
Cross-section, formula (18)					0·15	0·12	0·11

The behaviour of the total cross-section for velocities approaching that of light was found by using the formula (10) with the value of V given by (16), for the latter formula gives the small-angle scattering fairly well on the whole. If use is made of the fact that the second term in (16) contributes little at small angles, it follows approximately that

$$f = I/(1-v^2/c^2)^{\frac{1}{2}} a_0,$$

where

$$I = \int_0^\infty 2Z_p K^{-1} \sin Kr \, dr.$$

I can be found by numerical integration: its value is required only for a few suitably chosen values of K. The total cross-section is then given to the same degree of approximation by

$$Q = 2\pi \int_0^{2k} I^2 K\, dK/\{k^2 a_0^2 (1 - v^2/c^2)\}, \tag{18}$$

the integration of I^2K being carried out numerically using the values of I previously obtained.

For values of ka_0 above 100, the value of the integrand in (18) approaches the value 1000 atomic units for gold, and therefore as v tends to c, the elastic cross-section tends to $(2000/137{\cdot}2^2)\pi a_0^2 = 0{\cdot}11\pi a_0^2$.

Polarization by double scattering

Having calculated f and g for single scattering, the polarization produced by a second scattering through the same angle is given with little further calculation from the formula (5) As far as the scattering at 90° is concerned, Mott's conclusion that at high energies the asymmetry in double scattering for the Coulomb field of gold reaches a maximum of about 0·16 at about 120 keV is confirmed by Bartlett & Watson's calculations for the Coulomb field of mercury. In Paper A similar results were also reached for the atomic field of gold, and it was also shown that at certain energies below 1 keV large values of the asymmetry could occur, usually at those energies for which there was a pronounced diffraction minimum in the angular distribution.

Values of the asymmetry for angles other than 90° are given in table 3. It will be seen that even at high energies large polarizations can occur at angles other than 90°. At 5·4 and 16·5 keV, the largest values occur near a minimum in the angular distribution as will be shown by reference to figure 1 The reason is, as explained fully in Paper A, that the value of f is small and comparable with g near a minimum value of the scattering.

Table 3. Values of the asymmetry in double scattering for the atomic field of gold

energy keV	45°	60°	75°	90°	105°	120°	135°	150°	165°
5·4	0·00	0 02	0·00	0·05	0·03	0·01	0·04	0·02	0·00
16·5	0 00	0 05	0 01	0·01	0·00	0·37	0·07	0 03	0·01
33	0 00	0·01	0 00	0 05	0 09	0 14	0 13	0 04	0 00
121	0·00	0·01	0·02	0 14	0 3	0·4	0 5	0 4	0 3
392	0·10	0·01	0 00	0 16	0 7	0·9	1 4	1 9	0·1
1060	0 01	0·00	0 03	0 0	0 4	0 3	0·8	0 9	0 0

At 392 and 1060 keV large values of the asymmetry occur at angles greater than 90°. This remarkable result seems to be confirmed by the results of Bartlett & Welton for the atomic field of mercury. For 100 kV electrons they obtain values of the asymmetry which increase from 0·05 at 90° to a maximum value of 0·12 at 135°. For

230 kV electrons they give explicitly only the value at 90°, viz. 0·025, but their values of f and g lead to a maximum asymmetry of 0·6 at 150°.

The explanation of such large values is that for increasingly high energies and at large angles, the scattering—and therefore the value of f—becomes smaller and smaller. The value of g, which depends largely on the slowly varying quantities η_1 and η_{-2}, varies far less with energy. The result is that g and f eventually become comparable and consequently large polarizations can be expected where the signs of the real and imaginary parts of f are favourable to such a result. g of course vanishes at 180°, so that large polarizations should occur in general near the middle of the angular range, 90–180°. Since, however, the calculated value of f is least accurate in this region because it is small, the values of the larger polarizations given in table 3 are probably not very accurate. In any case they may depend somewhat on the form of the field at the larger distances, for the same reason that the angular distributions at these energies possess minima instead of falling monotonically as they would for a Coulomb field.

Further experiments should be carried out at angles other than 90° to test the above result, as the modifications of apparatus required for this purpose should not be serious There is at the moment, however, some doubt even as to whether a discrepancy with experiment still exists for the polarization at 90°. The recent experiments of Shull (1942) indicate that the observed polarization is, for some reason not yet understood, different according as to whether the electrons are scattered by the gold scattering foils through the 'reflexion' side of the foil or the transmission side. For the former, Dymond and Shull both obtain very small polarizations, while for the latter Shull obtains an asymmetry for 400 kV electrons in good agreement with the theoretical value.

Effect of modifications of field on scattering and polarization

Although the experimental evidence is not finally conclusive on the point, it may be that for certain energies the scattering and the polarization are definitely much less than predicted by theory. It is therefore of interest to consider to what extent modifications in the atomic field will affect the intensity of scattering and the magnitude of the polarization simultaneously. An atomic field of the form $Z_p e^2(1-e^{-\lambda r})/r$ was therefore taken for consideration, and the case of $ka_0 = 100$, $\lambda a_0 = 137$ investigated. The effect of adding on the repulsive field was to reduce δ_0 from 3·89 to 3·44, but δ_1 and higher phases were not appreciably reduced. On the other hand, $\frac{1}{2}(\eta_l - \eta_{-l-1})$ was reduced from 0·17 to 0·05. The reduction in δ_0 resulted in a reduction of the scattering at 90° to 0·4 of the previous value, and the reduction in $\frac{1}{2}(\eta_l - \eta_{-l-1})$ caused the asymmetry at 90° to be reduced to about 0·1 of the previous value. The reduction of the scattering at 180° was more marked, viz. to 0·1 of the previous value, while that at small angles was not much diminished.

The calculations seem to indicate in general that if the field is modified to reduce the scattering at 90°, the reduction will be markedly greater at larger angles, while

the asymmetry in double scattering at 90° will be reduced by a greater amount than the scattering. The values of λ necessary to produce appreciable reduction, however, involve modifications of the atomic field at distances which seem impossibly large. Thus taking $\lambda a_0 = 137$ reduces the atomic field to half its value at a nuclear distance of $2{\cdot}6 \times 10^{-11}$ cm., while the field is appreciably modified at still larger distances.

References

Bartlett & Watson 1940 *Proc Amer. Acad. Arts Sci.* **74**, 53.
Bartlett & Welton 1941 *Phys. Rev.* **59**, 281.
Champion & Barber 1938 *Proc Roy Soc.* A, **168**, 159.
Dymond 1934 *Proc Roy. Soc* A, **145**, 657.
Gupta 1939 *Proc. Phys. Soc.* **51**, 355.
Henneberg 1933 *Z. Phys.* **83**, 555
Klarmann & Bothe 1936 *Z. Phys.* **101**, 489.
Massey & Mohr 1941 *Proc Roy Soc.* A, **177**, 341.
Mott 1929 *Proc. Roy Soc.* A, **124**, 425
Mott 1932 *Proc. Roy. Soc.* A, **135**, 429.
Mott & Massey 1933 *Theory of atomic collisions.* Oxford. Clarendon Press.
Richter 1937 *Ann. Phys., Lpz*, **28**, 533.
Shull 1942 *Phys. Rev.* **61**, 198

The flame spectrum of carbon monoxide
III. The cool flame

By A. G. Gaydon, D.Sc.

Chemical Engineering Department, Imperial College, London, S.W.7

(*Communicated by Sir Alfred Egerton, Sec R S.—Received 3 March* 1943)

[Plates 3, 4]

The spectra of the cool flames, or pre-ignition glows, of carbon monoxide with oxygen and with nitrous oxide have been photographed The cool flame shows the same faintly banded spectrum as the normal flame, but this band structure is more clearly developed The OH bands are absent from the cool flame, which, however, shows strong sodium emission Cuprous chloride appears very readily as an impurity and the band systems of CuCl show a markedly different intensity distribution in the cool flames with oxygen and with nitrous oxide The application of the results to the theory of the combustion mechanism is briefly discussed

Introduction

The normal flame spectrum of carbon monoxide in the visible and near ultra-violet regions of the spectrum (Gaydon 1940, 1941) and in the photographic infra-red (Gaydon 1942*a*) has already been described, and its bearing on the combustion

problems has been discussed. It has been suggested that in the formation of normal carbon dioxide from normal carbon monoxide and oxygen an electronic rearrangement must occur at some stage, and that this will result in the newly formed molecules being initially endowed with considerable vibrational energy, the long life of which can account for the abnormal dissociation in the combustion products and the phenomenon known as afterburning.

It is known that a mixture of carbon monoxide and oxygen at a temperature just below the ignition point shows a 'cool flame', or more correctly a pre-ignition glow. The existence of this luminescence at temperatures below which it could arise from thermal excitation supports (Gaydon 1942*b*) the view that an electronic rearrangement must always take place during the combustion process. The conditions for the occurrence of this cool flame have been described by Prettre & Laffitte (1929) and Topley (1930), the colour being recorded respectively as *rouge violacé* and violet. The spectrum of the cool flame does not appear to have been recorded hitherto The normal flame of carbon monoxide is blue and the difference in colour would at first sight suggest a different spectrum However, in the present series of experiments the colour usually appeared to the author to be blue. The description of the colour of a very weak glow is rather uncertain, and the red glow from the walls of the furnace and the admixture of some orange sodium coloration with the true colour of the cool flame could no doubt produce a violet effect, and the colour of the cool flame obtained by the author was thus described on at least one occasion by Sir Alfred Egerton

In addition to the examination of the spectrum of the cool flame of carbon monoxide reacting with oxygen, some observations have also been made on a cool flame of carbon monoxide with nitrous oxide, and on the effect of traces of cuprous chloride on the spectrum of these cool flames.

Experimental

The experimental set-up was of the simplest. It consisted of an electrically heated furnace with two inlets and an outlet. The gases, stored in cylinders, were led through flow-meters to the inlets at atmospheric pressure and mixed in the furnace and then passed along to the outlet at the other end For the earlier experiments the reaction vessel consisted of a mild steel tube with a platinum-rhodium thermocouple cemented in at one end, and a quartz window, held against the flat end of the steel tube by a collar, at the other end, this permitted the luminescence in the tube to be viewed end on. For later experiments in which it was desired to eliminate effects due to metal surfaces, an all quartz reaction vessel with a sealed-on window was used, the thermocouple being dispensed with. The reaction vessels used were of 1 cm. internal bore and from 20 to 30 cm. long.

The gas flow was of the order 100 c c./min , and apart from a few brief visual observations it was found convenient to work with either fairly rich or fairly weak mixtures, usually $CO.O_2 = 4:1$ or $1:4$, as with anything approaching a theoretical

mixture it was difficult to maintain the furnace temperature sufficiently constant over long periods to give a reasonably intense glow without risk of ignition; on one occasion the resulting explosion shattered the quartz reaction vessel, and on several occasions exposures were spoilt by the gases igniting.

Spectra were recorded on a small quartz spectrograph (Hilger E.3) and on a large aperture glass prism spectrograph giving a dispersion of about 50 A/mm. in the blue-green With both instruments exposure times varied from ½ to 6 hr. according to slit width and intensity of the glow. A neon glow lamp containing some hydrogen and helium was used as a comparison spectrum, as, with the small dispersion and wide slit usually used, the iron arc spectrum was difficult to recognize.

The cool flame with oxygen

Mixtures of carbon monoxide and oxygen showed a blue luminescence when passed through the furnace at a temperature of about 600° C. The glow appeared strongest along the axis of the reaction vessel and the volume within about 1 mm. of the walls appeared dark In some of the later observations with the steel reaction vessel, after it had become slightly rusted, it was noted that specks of dust or rust sometimes glowed brightly, and in some cases became hot enough to cause ignition, this suggested that, in the steel vessel, surface catalysis was playing a part in the reaction.

The spectrum was successfully photographed on both the quartz and glass spectrographs Spectrograms of both the normal and cool flames are reproduced in strips *a* to *d* of plate 3. It will be seen that the normal and cool flame spectra are essentially similar, but with certain significant differences

The normal flame shows a faintly banded spectrum on a relatively strong continuous background, being strongest in the blue and near ultra-violet With the small dispersion and rather wide slit used in the present experiments, the banded structure does not show up very well in the normal flame. In the cool flame, however, using the same slit width and instrumental adjustments, the band structure is much more obvious. Strip *b* of the plate shows a fairly regular development of this band structure in the region 3200–3400 A, and strip *d* shows the greater clarity of the banded structure in the visible region Thus the spectrum of the cool flame resembles more nearly that of the afterglow of carbon dioxide in a discharge tube (Fowler & Gaydon 1933) or the normal flame at reduced pressure This sharpening of the band structure in the cool flame and in the flame at reduced pressure probably results from the lower rotational temperature of the molecules, the branches of the complex rotational structure being shorter so that the heads of the bands stand out more clearly against the general background of unresolved structure.

Another difference between the spectra of the cool and normal flames is the absence of the OH band at 3064 A from the former. This band is extremely per-

sistent in the spectrum of the normal flame and can only be eliminated by the most careful drying. Most of the work on the cool flame was done with moderately dry gases, no special drying was undertaken, but the gases were stored in cylinders at around 100 atm. pressure, so that the gases were presumably not more than 1 % saturated at room temperature. To confirm the absence of the OH band in the cool flame, very long exposures were given with gases deliberately saturated with moisture at room temperature, but no trace of the band was found. The absence of the OH band forms the subject of later discussion.

A third difference between the spectra of the normal and cool flames was the strong excitation of the sodium D lines in the cool flame. This is well shown in strips *b* and *d* of the plate The sodium emission was initially so strong in the cool flame that it produced a yellow coloration of the whole flame, and it was only after the flame had been running for an hour or so that it died down sufficiently for the blue colour to become predominant. It is clear that the strong excitation of the D lines at a temperature of around 600° C must be due to the atoms in some way taking up energy directly from the combustion process, and cannot be thermal in origin. This effect leads the author to doubt the reliability of the sodium-line reversal method of measuring flame temperatures, for which thermal excitation of the sodium must be assumed.

The strong excitation of the sodium resonance lines in the cool flame suggested that it might be possible to arrive at an estimate of the amount of energy available for the excitation by observing the resonance lines of other metals. Accordingly a little metallic cadmium, which boils at 767° C, was placed in the reaction vessel, and a long exposure was taken on the cool flame using special ultra-violet sensitive plates; no sign of the Cd resonance line at 2288 A was observed. Similarly, placing a little zinc-mercury amalgam in the tube failed to bring up either the Zn line 2138 A or Hg 2537 A. Under conditions in which the CuCl bands were strong (see later section) no sign of the Cu resonance lines at 3274 and 3248 A appeared. Negative evidence of this type is never very convincing as it is difficult to be sure that an adequate concentration of atomic vapour was present, it can merely be said that no positive evidence was obtained to indicate the existence in the reacting gases of any atom or molecule with a metastable electronic state of high energy.

Some of the plates taken seemed to indicate that the emission from the cool flame, relative to that from the normal flame, was rather stronger at the red end and less strong at the short wave-length end. This effect could not, however, be definitely confirmed owing to the chromatic aberration of the quartz condensing lens used to throw an image of the cool flame on to the slit of the spectrograph; with a relatively small source it was found that the strength of the spectrum in any region was markedly affected by the adjustment of this lens, it being only possible to focus the image of the cool flame accurately on to the slit for one wave-length. The cool flame extended throughout the length of the reaction vessel, while the normal flame produced when the gases were allowed to ignite was situated at the far end of the tube. Nevertheless, it would not be surprising if at

the lower temperature of the cool flame the vibrational intensity distribution was slightly modified in such a way as to strengthen the red and weaken the ultra-violet part of the spectrum. The strengthening of the red end shown in strip *d* of the plate is on the original negative and is not accentuated by the process of enlarging, but may be due, at least in part, to the effect of the condensing lens.

The cool flame with nitrous oxide

It was found that the reaction between carbon monoxide and nitrous oxide below the ignition temperature was also accompanied by luminescence. As first obtained, using nitrous oxide of anaesthetic quality in a steel reaction vessel, the cool flame was green; the green colour was especially clear for mixtures rich in CO The spectrum under these conditions showed a well-marked band system which was, thanks to a suggestion by Dr R W B. Pearse, quickly identified with copper chloride, CuCl These bands are discussed in detail in the next section.

Using a quartz reaction vessel the CuCl bands were less troublesome, although they still occurred faintly on occasions, especially after an explosion in the tube had stirred up dust in the rubber connexions, etc. The spectrum of the pure CO-N_2O cool flame was photographed on the quartz spectrograph and some visual observations were made with the large aperture instrument Apart from the absence of the OH band there appeared to be little difference between the cool and the normal CO-N_2O flame. It has recently (Gaydon 1942*a*) been pointed out that the normal CO-N_2O flame differs slightly from the CO-O_2 flame in showing in addition to the blue faintly banded structure a continuous emission in the yellow-green. This continuum in the yellow-green was probably also present in the spectrum of the cool flame, especially when the gases were just on the point of ignition, the cool flame changing from a bluish white to a more yellow shade as the temperature of the gases rose to the ignition point The identification of a continuous emission is not easy owing to the absence of distinctive features for measurement, but the region of the spectrum and colour sensation produced by the emission of this continuum suggests a possible identity with continuous emission characteristic of the air or so-called oxygen afterglow. There is evidence to show that the air afterglow continuum results from a reaction between nitric oxide and atomic oxygen (Spealman & Rodebush 1935). If the two continua are the same this would indicate that atomic oxygen was present at any rate in the normal CO-N_2O flame, and probable just before ignition in the cool flame. The author has previously (Gaydon 1942*b*) been rather sceptical of combustion mechanisms involving atomic oxygen, owing to the failure of carbon monoxide to give an atomic cool flame with atomic oxygen (Harteck & Kopsch 1928). It is hoped to follow up the possible identity of the flame continuum with the afterglow continuum with a view to using it as a test for atomic oxygen in flames. The sodium lines were less persistent in the cool flame with nitrous oxide than with oxygen.

CuCl BANDS IN THE COOL FLAMES

Strong CuCl bands were observed in the spectrum of the cool flame of nitrous oxide and carbon monoxide when using the steel reaction vessels. They did not appear, except faintly at times, in the quartz reaction vessel Also they did not appear with the cool flame maintained by oxygen even in the steel tube unless a chlorine compound (HCl vapour) was added to the carbon monoxide, when they appeared very strongly. It thus appears that the copper must have been present in the steel tube, while the chlorine must have been carried by the nitrous oxide The amount of chlorine in the nitrous oxide, which was of anaesthetic quality, must have been extremely small, it was possibly carried in as ammonium chloride, which is rather volatile.

The spectrum of CuCl consists of five band systems, usually known as A, B, C, D, and E, with (0, 0) bands at $\lambda\lambda$ 5262, 4881, 4847, 4354 and 4333 respectively. In absorption all the systems appear fairly readily, but in emission in a flame or arc the D and E systems are usually dominant. Strip *e* of plate 4 shows the CuCl bands as obtained when cuprous chloride is introduced into an ordinary Bunsen flame. Systems D and E are strong and show flutings due to other bands in the sequences, while systems B and C are less marked The green region is covered by the strong patch of continuous or nearly continuous emission characteristic of all flames containing copper, the origin of this is uncertain, but may be due to an oxide of copper When cuprous chloride is introduced into an electric arc systems D and E are strongly developed, but systems B and C are weak and A is absent (see plate published by Pearse & Gaydon 1941).

When obtained in the cool flame with oxygen, the intensity distribution of the CuCl bands is similar to that in the ordinary flame or the arc, system A being absent, and systems B and C weak compared to D and E This is shown in strip *g* of the plate The bands in the cool flame do not show the flutings as they do in the hot flame, and the (1, 0) band of system D is rather less intense This is presumably due to the lower vibrational temperature of the CuCl molecules in the cool flame. The effect is real, and not due to instrumental differences, as the plates were taken with the same slit widths and adjustments.

With the cool flame maintained by nitrous oxide (strip *f*, plate 4) there is a very marked difference in the intensity distribution among the CuCl systems. Systems B and C are now relatively stronger compared with D and E, and the (0, 0) and (0, 1) bands of system A occur strongly, the apparently greater strength of the (0, 1) band is chiefly due to the higher plate sensitivity in this region. In strip *f* these bands are well shown. The patch of continuum at the red end is due to thermal emission from the walls of the furnace, but the marked strengthening of system A is not caused by this thermal emission; the results have been confirmed by visual observation. Also, the cool nitrous oxide flame containing CuCl is green in colour, while the cool flame with oxygen and CuCl is blue.

Carbon monoxide and cuprous chloride have a strong chemical affinity, and

indeed a solution of cuprous chloride is the usual reagent for absorbing carbon monoxide. In solution a complex, probably $CuCl.2H_2O.CO$, is formed. It seems probable to the author that some similar complex is formed in the gas phase and that this reacts more readily with oxygen or nitrous oxide than does carbon monoxide itself. The CuCl thus acts as a catalyst, and some of the heat of the reaction is unloaded on to the CuCl molecules to produce the emission spectrum of this molecule. Combustion with nitrous oxide should liberate more heat than combustion with oxygen, but it is clear, nevertheless, that the CuCl molecules are more highly excited by the reaction with O_2 than with N_2O. This suggests that the reaction mechanism is not simple It is well known that the blue flame produced by throwing salt (NaCl) on a coal fire shows CuCl bands, this may be another example of catalytic oxidation of carbon monoxide by cuprous chloride.

The presence of cuprous chloride seems to suppress the sodium lines in the spectrum of the cool flame. See strips *f* and *g* of plate 4

Discussion

The emission of the normal carbon monoxide flame spectrum by reacting gases at a temperature of around 600° C strongly supports the view previously put forward that an electronic rearrangement of the CO_2 molecules must occur following the initial combustion process.

There is not much evidence to suggest that any high energy of excitation is freely unloaded on to any atom or molecule with suitable energy levels which may be present Thus, apart from the failure to excite spectra of Cd, Zn or Hg, certain other impurities such as FeO and FeCl should surely have been introduced by the walls of the steel tube more readily than CuCl, yet fail to show in the spectrum. In agreement with direct combustion experiments, it seems that carbon monoxide and oxygen, or nitrous oxide, are reluctant to react directly, but do so more readily in the presence of a catalyst. The intervention of the most minute amounts of CuCl has already been discussed It has been shown by Magee & Ri (1941) that the reaction $H + H + Na = H_2 + Na^*$ occurs with fairly high efficiency when sodium is introduced into atomic hydrogen. It is possible that the excitation of the sodium lines in the cool flame is the result of a different but similar type of reaction.

One of the most significant differences between the spectrum of the cool and normal flames is the absence of the OH band from the former. Topley (1930) has shown that the reaction rate under cool flame conditions is approximately proportional to the concentration of moisture, and he discusses the possibility of the reaction being maintained by a chain mechanism in which either H_2, H_2O_2 or OH take part. The absence of the OH band indicates absence of electronically excited OH radicals in the cool flame, but not of course absence of unexcited OH. However, the absence of the band leads one to suspect that the catalytic action of water on the cool flame is not due to the maintenance of the usual water-gas equilibrium, but is more likely to be connected with the action of the H_2O molecules

in quickly setting free the vibrational energy stored in the newly formed CO_2 molecules, and so assisting the thermal maintenance of the reaction.

In conclusion, it is a pleasure to thank Sir Alfred Egerton once again for his keen interest in the work. I am also much indebted to the Council of the Royal Society for financial assistance.

References

Fowler, A. & Gaydon, A. G. 1933 *Proc. Roy. Soc.* A, **142**, 363.
Gaydon, A. G. 1940 *Proc. Roy. Soc.* A, **176**, 505.
Gaydon, A. G 1941 *Proc. Roy. Soc.* A, **178**, 61.
Gaydon, A. G. 1942*a* *Proc. Roy. Soc.* A, **181**, 197.
Gaydon, A. G 1942*b* *Spectroscopy and Combustion Theory.* London: Chapman and Hall.
Harteck, P & Kopsch, U. 1928 *Z. phys Chem* **12**B, 327.
Magee, J. L. & Ri, T. 1941 *J. Chem Phys.* **9**, 638.
Pearse, R. W. B. & Gaydon, A. G. 1941 *The Identification of Molecular Spectra.* London Chapman and Hall.
Prettre, M. & Laffitte, P. 1929 *C.R. Acad. Sci., Paris*, **189**, 177.
Spealman, M. L. & Rodebush, W. H 1935 *J. Amer. Chem. Soc.* **57**, 1474.
Topley, B. 1930 *Nature, Lond*, **125**, 560.

Description of Plates 3, 4

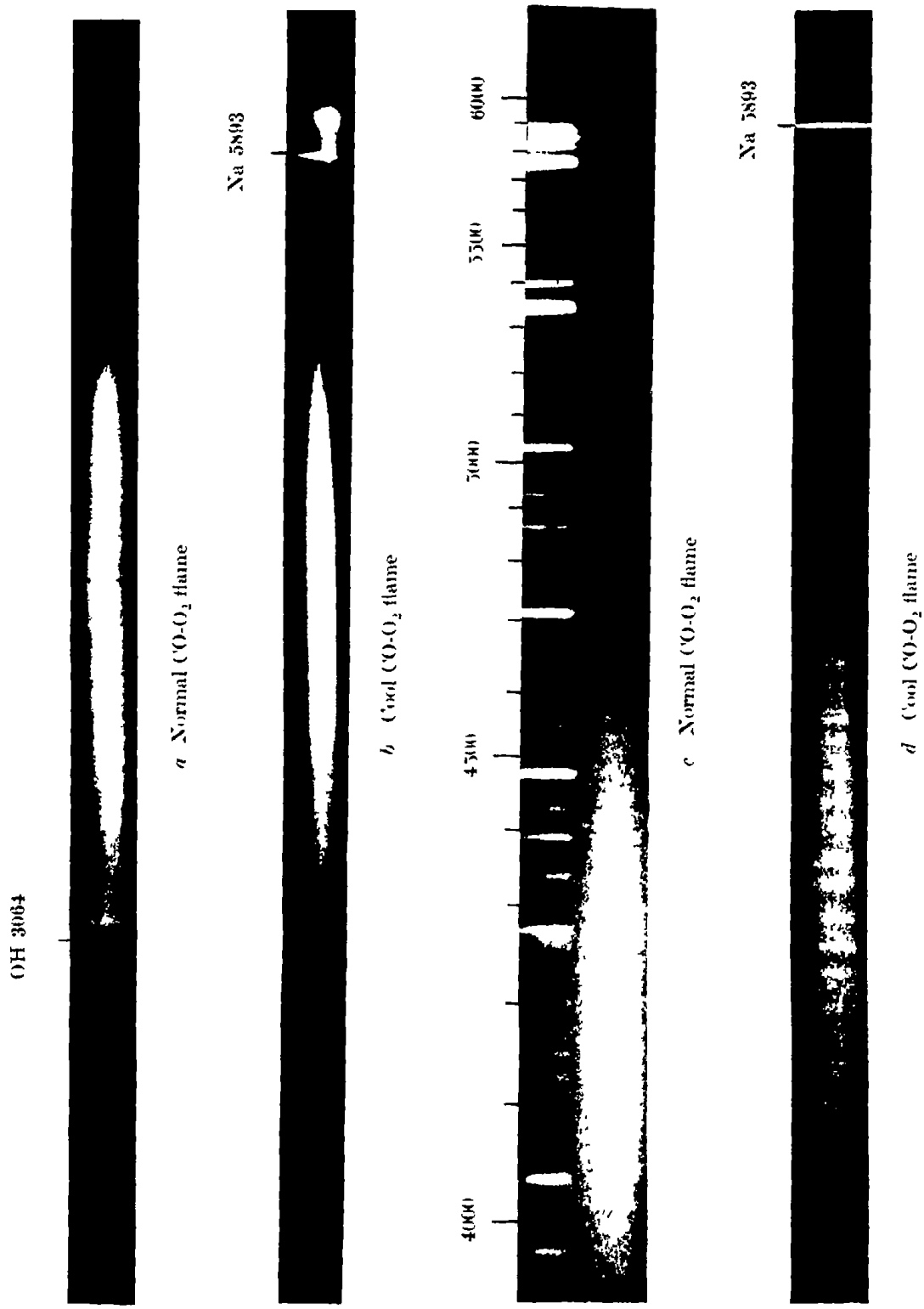
Strips *a* and *b* are enlargements of spectra taken with the small quartz spectrograph using Ilford Process Panchromatic plates. Strips *c* to *g* are of spectra taken on the large aperture glass instrument using Ilford Press Ortho (series 2) plates.

Plate 3

a. Normal flame of moist carbon monoxide burning in oxygen. 4 min.
b. Cool flame of carbon monoxide and oxygen. Furnace temperature 610° C. $CO.O_2$ flow = 1:4. Same slit width and adjustment as *a*. Note the band structure at the short-wave end and the sodium line. The continuum at the red end is from the walls of the furnace. Exposure 6 hr.
c. Normal flame of CO burning in O_2. 2 min. Neon (with H and He) comparison above.
d. Cool flame CO and O_2. 3 hr. Same slit width and adjustment as *c*.

Plate 4

e. Cuprous chloride in Bunsen flame. 4 min. Neon comparison above.
f. Cool flame of CO and N_2O in steel reaction vessel showing CuCl bands. Same slit width and adjustment as *e*. The (0, 0) bands of the CuCl systems are indicated. Exposure 1 hr. 15 min.
g. Cool flame of CO, which had been bubbled through HCl, and O_2 in steel reaction vessel showing CuCl bands. Same slit width and adjustment as *e* and *f*. Exposure 2 hr.



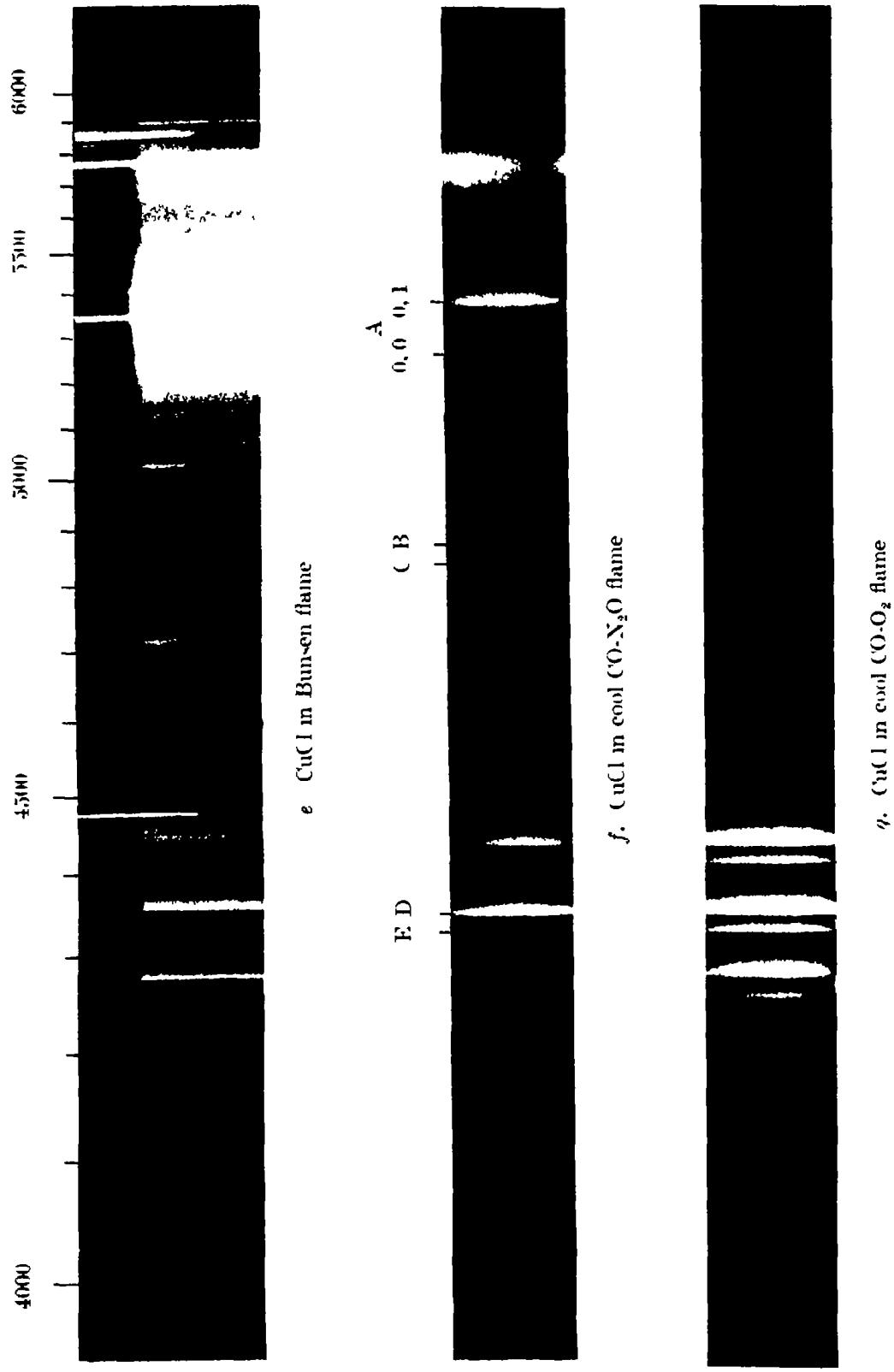
a Normal CO-O_2 flame

b Cool CO-O_2 flame

c Normal CO-O_2 flame

d Cool CO-O_2 flame

(*Facing p* 206)



e. CuCl in Bunsen flame

f. CuCl in cool CO-N_2O flame

g. CuCl in cool CO-O_2 flame

The refraction and dispersion of sulphuretted hydrogen, sulphur dioxide and carbon oxysulphide

By H. Huxley, M.Sc., A.Inst.P. and H. Lowery, D.Sc., F.Inst.P.
South-West Essex Technical College, Walthamstow, E.17

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 30 *March* 1943)

[Plate 5]

An advance in the technique of using the Jamin interferometer for the determination of the dispersion of gases has been effected by making a systematic allowance for fringe-drift by means of photographic measurements. New values of the refraction and dispersion of sulphur dioxide and sulphuretted hydrogen have been determined, reduced to 'infinite dilution' by means of newly measured pressure coefficients, in order to allow for departure from the gas laws. The refraction and dispersion of carbon oxysulphide have been measured for the first time.

Introduction

The primary object of this investigation was to obtain the refraction and dispersion of carbon oxysulphide which have not previously been measured. A Jamin interferometer was used for this purpose, as in earlier experiments (Lowery 1931); certain improvements in technique were contemplated, hence it was decided to repeat measurements on sulphuretted hydrogen and sulphur dioxide, especially in view of the high probable error in the existing published values relating to these substances. At the same time their dispersion curves were extended downwards from 5000 to 4000 A.

General discussion

(*a*) *Limiting values*

The values of gaseous refractivities as determined by different experimenters are notoriously at variance, due largely to the mode of reduction of the results; this is true even where the workers have used a reliable experimental technique, and where the substances employed have been of the highest degree of purity (though sometimes discrepancies have been wrongly ascribed to the want of purity of the specimens instead of to the process of reduction). This may be seen even in the vast amount of work that has been done on such an easily manageable substance as air (Cuthbertson 1910*a*).

The bulk of refractivity data has been reduced to 'limiting values' at N.T.P., thus involving two fundamental assumptions, viz. (*a*) the relationship between refractivity and gas density, and (*b*) the relation between gas density and pressure and temperature.

Some attempts have previously been made to reduce refractivity data through direct density determinations, e.g. by incorporating a density bulb in the apparatus, but the errors introduced due to the weighing process, since very small differences are involved, are greater than those from pressure and temperature measurement, thus Cuthbertson (1910*b*) in using both the density and pressure/temperature methods for SO_2, showed a *probable error* of ± 21 in 7000 in the former, as compared with ± 7 in 7000 for the latter. Of course, in some cases the density method is the only method available, but experience seems to indicate that where pressure/temperature measurements are feasible, the procedure should be dictated thereby.

Gladstone and Dale's law, $(n-1)/d$ = constant, may be taken as applicable for the pressures (below 2 atm) used so far in most refractivity experiments. Much work, however, proves valueless through simple reduction by the perfect gas equation, $pv = RT$, without any indication as to the ranges of pressure and temperature of the experiments, and it is time that new issues of collections of physical tables should omit the 'indefinite' values of Arago, Biot and Dulong, except perhaps in the extremely rare cases where no revised modern values are available. The minimum information given in any such recorded data must include the mean pressure and temperature, and the actual relationships assumed in the reduction to N T P It is remarkable that reduction procedure is still in such an unsatisfactory state, seeing that the essence of its importance was realized by the pioneers, Mascart and Ketteler (*c* 1870), not to speak of later workers. These earlier experimenters dealt with the reduction in two parts, viz. (*a*) pressure. $(n-1) \propto p(1+ap)$ and (*b*) temperature: $(n-1) \propto 1/(1+bt)$, where the pressure and temperature coefficients, a and b respectively, were determined by experiment.

Cuthbertson (1910*b*) discussed his results for sulphur dioxide and sulphuretted hydrogen by using the ratio (theoretical density)/(experimental density) Fajans, Wust & Reindel (1934), working on sulphuretted hydrogen, used van der Waal's equation to reduce to N T P. conditions, and then Cuthbertson's ratio in the discussion of the results. The use of van der Waal's equation is only an approximation, since the constants used are somewhat artificial.

In the present work, the reduced values of refractivity were obtained from the relation

$$(n-1)_{\text{N T P.}} = N\lambda/l \,.\, 760/p\{1+(t/273{\cdot}18)\},$$

where N is the number of interference fringes counted, l is the length of the refraction tubes, and λ the wave-length of the light. Values of $(n-1)_{\text{N T P.}}$ were found at a series of pressures between 400 and 760 mm. from which a graph of $(n-1)_{\text{N T P.}}$ against pressure was drawn in order to determine the coefficient a. In all cases these graphs were 'good' straight lines, so that by extrapolation the values of $(n-1)_{\text{N T P}}$ at 0 and 760 mm. pressure could be determined, 'a' being calculated from

$$(n-1)_{\text{N T P}} \text{ at } 760\,\text{mm.} = (n-1)_{\text{N.T.P.}} \text{ at } 0\,\text{mm.} \times (1+ap),$$

where $p = 760$ mm. Thus $(n-1)_{\text{N T P}}$ at 0 mm. is a corrected value of refractivity for conditions under which the gas may be regarded as behaving according to the laws

of a perfect gas, i.e. when the gas is at *infinite dilution*. This value of refractivity will be represented by $(n-1)_0$. The value of $(n-1)_{pt}$ for a given pressure and temperature may be found from

$$(n-1)_{pt} = (n-1)_0\{p(1+ap)/760)\}\{(1+bt)/(1+t/273{\cdot}18)\}.$$

The values of a and b will be those which have been determined experimentally as mentioned above for each individual gas, and comparison between the work of various investigators will only be possible when all their results are reduced in the same way to *infinite dilution*.

(*b*) *Fringe drift*

Another source of confusion in recorded values of gaseous refractivities arises from a slight movement of the interference fringes across the fiduciary line of the telescope even when pressures and temperatures are kept constant. This movement is known as 'fringe drift', and has been noted by nearly all who have made measurements, though very little attempt has been made to eliminate its effects or correct for its presence, except in the work of C. and M. Cuthbertson on neon and helium (1932) in which the drift due to temperature changes in the apparatus was allowed for. Where experimenters give several values of refractivity at a given wave-length, it will usually be noted that considerable variations from the mean value occur, variations so great that they should not be expected in work with such a sensitive instrument as an interferometer. These are undoubtedly due to the presence of drift, which leads to an inaccurate value for N in the above expression for refractivity.

Using the probable error formula

$$E = 0{\cdot}6745\sqrt{[\Sigma e^2/(q-1)]},$$

where q denotes the number of observations and Σe^2 the sum of the squares of residuals, the following calculated values of probable error are obtained, which can hardly be regarded as satisfactory.

(1) SO_2: Cuthbertson (1908). Density bulb method, $E = \pm 21$ in 7000, Pressure/temperature method, $E = \pm 7$ in 7000. Maximum variation shown is 56 in 7000, equivalent to one whole fringe movement.

(2) H_2S: Fajans *et al.* (1934) $E = \pm 7$ in 6000. This result is in spite of the fact that thirty observations were made around the same pressure (about 230 mm.). Maximum variation shown is 29 in 6000, which is equivalent to one whole fringe movement.

(3) *Air*: Meggers & Peters (1918). Their own estimated probable error is $\pm 3{\cdot}8$ in 3000, which again can scarcely be regarded as entirely satisfactory, particularly as the work is regarded as 'classical', being apparently regularly used in computing astronomical data. Similarly, the probable error in a determination of air by one of the present authors (Lowery 1927), using the Jamin interferometer by the visual method, is ± 3 in 3000.

In order to achieve the accuracy necessary for consistent results, the fringe displacement due to interposing a gas column in one of the beams of the Jamin interferometer must be measured to 1/50th of a fringe in a minimum number of 250 fringes. This accuracy of observation cannot be obtained with visual observations, especially with the drift effect present.

Drift may be (*a*) permanent, (*b*) temporary. (*a*) Permanent drift is due to changing external conditions and may affect the value of N positively or negatively. In the present work it was reduced to negligible proportions by heat insulation of all parts of the apparatus and careful control of room temperature. (*b*) Temporary drift is due to change in gas density in the refraction tube, and will, of course, disappear when the gas is removed. With gas in one tube, the drift was always negative, of the order 0·5 fringe per hour, the value depending on the total pressure of the gas column. This indicates a decrease in gas density, presumably due to absorption by the glass apparatus and the mercury surface in the manometer. After evacuation, this drift became positive.

Walker (1903), in his work on SO_2, mentions this drift, stating that it brought his accuracy of fringe measurement down to one-tenth of a fringe, but he did not make any attempt to allow for it in his calculations Other experimenters have recorded their elaborate precautions to eliminate the drift, but have really only prevented permanent drift. A possible exception may occur in the case of perfectly dry air and with gases like hydrogen and oxygen far from their liquefying points. Sears & Barrell (1939) make no mention of drift in their work on air

As a result of the application of photographic observations to the Jamin interferometer as shown below, the probable errors for the gases measured were as follows, and indicate a marked improvement in the use of this type of interferometer:

Air	± 0·2 in 3000	H_2S	± 0·5 in 6000
SO_2	± 1·7 in 7000	COS	± 0·4 in 9000

Experimental

A Jamin interferometer with Hilger constant deviation monochromatic illuminator was mounted with suitable apparatus for the admission of the gas under test in one beam of the interferometer under known conditions of temperature and pressure. An Osram pearl 100 W lamp was used to obtain a continuous spectrum when necessary, and a mercury-cadmium vapour lamp as a source of spectral lines.

In the visual method of using the Jamin interferometer, light from a monochromatic illuminator passes through a collimator which directs a parallel beam through the mirrors. This beam is finally received by a telescope in which interference phenomena are observed due to optical differences in the paths between the mirrors. For the purpose of photographing the interference fringes, the above arrangement was modified as follows: A parallel beam of white light was directed through the mirrors by means of a collimator. This was received by a Hilger constant deviation spectroscope provided with a specially made camera attachment, the

various spectral regions being brought into focus on the camera plate by means of an auxiliary lens in place of the usual eyepiece of the spectroscope. Preliminary focusing of the fringes in the camera was done visually with a ground-glass screen in place of the photographic plate. The various spectral lines on the plate were of course crossed by interference fringes as illustrated in figures 1 and 2 (plate 5).

Photographs of the fringe system were taken under measured conditions from which, knowing the approximate number of fringes involved due to a given gas density, the excess fraction of a fringe may be accurately determined. This latter is obtained from the displacement between the fringe systems for a given spectral line (*a*) with, and (*b*) without, the gas in the tube. The fringe width varied from about 2 mm. for yellow light to 1·2 mm. for violet light, and was measured accurately to 1/50th of a fringe by means of a vernier microscope.

The importance of drift measurement has been stressed above. To facilitate this measurement, pressure and temperature readings were always taken 5 min. before making an exposure

A line exposed over a known interval will only record half the actual drift during this interval, as it is the displacement of the band centre which is measured. To bring all fringes to the positions they had at the beginning of the exposure, the allowance is half the drift for the interval of the exposure. To each fraction must be added the drift during the 5 min interval between the pressure reading and the beginning of the exposure; this corrects the fringe position for all lines to the instant of pressure measurement.

The value of the drift was measured by exposing the mercury green line (10 sec.) at the beginning and again at the end of the full exposure Overlapping was avoided by slight rotation of the Hilger prism.

Such measured drift applies to the mercury line only; the drift for each wave-length will be inversely proportional to the band width for the line—the band width varies directly as wave-length. The band width, as measured on the plates, was taken in preference to the wave-length, so as to avoid any error in the focusing of each line on the plate.

In general the gas under test was liquefied or solidified in U-tubes and fractionated into a specially constructed pyrex glass bulb which was then attached to the gas apparatus of the interferometer. A long tube of phosphorus pentoxide was permanently connected to the apparatus for drying. Final traces of air were removed by means of a Hyvac pump with the substance still in the solid state. After use the gas was removed by means of soda-lime and granulated charcoal. Before experiment the gas was always admitted to the apparatus for 1 hr. in order that it should acquire a steady temperature, viz. that of the tank of water which surrounded the refraction tubes.

In order to obtain the pressure coefficient, *a*, gas was admitted to the apparatus progressively in amounts corresponding to approximately 50 fringes until the pressure was about one atmosphere, pressure and temperature measurements being taken between each step of the process.

No allowance for drift was made at this stage though the values of $(n-1)_{N.T.P.}$ thus obtained for mercury green light were utilized in estimating the number of fringes, N, in the measurement of the dispersion plates. The accurate value of a was obtained from the slope of the graph (= a straight line) obtained by plotting the values of $(n-1)_{N\,T\,P.}$, calculated from the dispersion plates, against their respective pressures which were within the range 400–760 mm.

The measurement of the temperature coefficient was not attempted as all readings for a given gas were made at approximately the same temperature. Assuming that the ratio (theoretical density/experimental density) covers both pressure and temperature variations from the general gas equation, the temperature coefficient may be estimated using the measured value of a.

Estimation of dispersion entails the determination of the approximate values of $(n-1)$ for each wave-length over the range The value of $(n-1)$ for the mercury green line was estimated from the approximate pressure coefficient curve and the number of fringes displacement for this line and each of the lines in question found by superimposing a system of interference due to a continuous spectrum over the normal plate exposure. After a displacement of 100–150 fringes from the achromatic fringe due to allowing a few cm. of gas into the refraction tube, the white light fringes become very close together and are difficult to photograph, hence the estimation plates have to be made at low gas pressure. The value of $(n-1)_{N\,T\,P}$ was calculated for each chosen wave-length and the ratio of the value for each line against that for the mercury green line found and checked by repeat experiments. The fractions so obtained were used for the time/dispersion plates of which half a dozen were taken for each gas at pressures of 45–70 cm. involving 300–600 fringes according to the nature of the gas. The values of $(n-1)$ thus obtained were reduced to the *infinite dilution* limit and used for drawing the dispersion curve.

(*a*) *Sulphuretted hydrogen* (H_2S)

Sulphuretted hydrogen was prepared by the action of hydrochloric acid on ferrous sulphide; after washing with water it was passed for several hours through a suspension of magnesium oxide in distilled water, thus forming magnesium hydrogen sulphide, $Mg(HS)_2$, which when warmed yielded the gas together with water vapour. The latter was removed by means of phosphorus pentoxide and the gas was liquefied and fractionated, using solid carbon dioxide and ether as the cooling agent, giving a product from which the gas could be obtained for refraction measurements.

The pressure coefficient a, for use in the relation

$$(n-1)_{N\,T\,P.} \text{ at } p\,\text{mm.} = (n-1)_{N\,T\,P.} \text{ at } 0\,\text{mm.} \times (1+ap)$$

was found experimentally to be $a = 0{\cdot}1087 \times 10^{-4}$ per mm.

Assuming the ratio (theoretical density/experimental density) to represent the total deviation from the perfect gas laws, the temperature coefficient, b, may be estimated using the value found for a. From the values theoretical density = 1·1747

(air = 1) and experimental density = 1·1895, the temperature coefficient $b = 0{\cdot}003914$ per ° C.

No previous values of the pressure and temperature coefficients have been determined.

Table 1 gives the values of the refractivity, reduced to infinite dilution, at various wave-lengths.

TABLE 1. DISPERSION OF SULPHURETTED HYDROGEN (H_2S)

λ	$(n-1)_0 \times 10^7$	λ	$(n-1)_0 \times 10^7$
5792·26	6408·3	4801·25	6519·8
5771·20	6409·9	4679 46	6539·2
5462·25	6437·9	4359·56	6599·4
5087·23	6479·8	4047·70	6674·6

The Sellmeier dispersion formula calculated from these results by the method of least squares is found to be

$$(n-1) = 4{\cdot}520 \times 10^{27}/(7322 \times 10^{27} - \nu^2),$$

where ν is the frequency of the light.

Remarkably few refractivity measurements have been made on this commonly occurring gas. The values given by Arago in 1805, Dulong in 1826, and Mascart in 1878, although frequently given in modern collections of tables (e g. Kaye & Laby 1941) are of historical interest only. The two modern determinations are by Cuthbertson (1910*b*), and by Fajans *et al.* (1934)

Cuthbertson, working at 760 mm. and 16° C, found $(n-1)_{N\,T\,P.} \times 10^7 = 6499{\cdot}3$ for the mercury green line Reduction of this value for comparison with the present measurements gives $(n-1)_0 \times 10^7 = 6439{\cdot}4$

Fajans *et al.* made thirty determinations at about 226 mm. and 20° C counting 200 fringes. Their value (reduced by means of the general gas equation and Gladstone and Dale's law) for the mercury green line is $(n-1) \times 10^7 = 6458$, which when reduced to infinite dilution gives $(n-1)_0 \times 10^7 = 6443{\cdot}6$; thus we have the values in table 2 for comparison.

TABLE 2. H_2S

λ 5462·25	$(n-1)_0 \times 10^7$	E
Fajans *et al.*	6443·6	± 7 in 6000
present work	6438	± 0 5 in 6000

Cuthbertson states that he took five measurements with a maximum variation from the mean of 1 in 300. The mean value only is stated, so that E cannot be calculated.

(*b*) *Sulphur dioxide* (SO_2)

The gas was obtained direct from a siphon, frozen in a U-tube by means of solid carbon dioxide, put under a high vacuum pump, and then admitted to the apparatus after passing over phosphorus pentoxide.

The refractivity results and other data obtained in the present work are as follows:

pressure coefficient	$a = 0{\cdot}2041 \times 10^{-4}$ per mm.
theoretical density	$= 2{\cdot}2123$ (air = 1)
experimental density	$= 2{\cdot}2639$
temperature coefficient	$b = 0{\cdot}004190$ per °C

TABLE 3. DISPERSION OF SULPHUR DIOXIDE (SO_2)

λ	$(n-1)_0 \times 10^7$	λ	$(n-1)_0 \times 10^7$
5792·26	6599·4	4801·25	6699·2
5771·20	6601·3	4679 46	6718·3
5462·25	6627·1	4359 56	6773·8
5087·23	6661·6	4047·70	6843·3

The Sellmeier formula corresponding to table 3 is

$$5{\cdot}220 \times 10^{27}/(8178 \times 10^{27} - \nu^2).$$

Measurements of the refractivity of sulphur dioxide have frequently been made in the past, no doubt owing to the comparative ease of obtaining the gas in a pure state. Previous values of the pressure and temperature coefficients have also been recorded, as shown in table 4.

TABLE 4. SO_2

	$a \times 10^4$ per mm.	b per °C
Mascart (1874)	0·25	0·00471
Mascart (1878)	0 25	0 00460
Ketteler (1885)	0·214	0·00411
Walker (1903)	0·398	0·00416
present work (1942)	0·2041	0 00419

Walker's (1903) coefficients, when combined, do not agree with the ratio (theoretical density)/(experimental density) and it is not clear whether he used some other form of reduction formula. For his refraction experiments he used sodium light and a mean pressure of 650 mm. and 16° C.

Cuthbertson (1908) worked at very low pressures, his maximum being less than 200 mm., and assumed the gas to be perfect under these conditions. For purposes of comparison his mean pressure is taken as 100 mm., the actual pressure not being stated.

Stuckert (1910) used a prism method and found $(n-1) \times 10^7 = 6666$ for the mercury green line, but his results cannot be reduced for comparison purposes owing to lack of precise data as to his experimental conditions.

Reducing the above-mentioned results where possible to infinite dilution, we have the following figures for comparison (table 5).

TABLE 5. SO_2

	$\lambda = 5462{\cdot}25$ $(n-1)_0 \times 10^7$	E
Walker (1903)	6633	± 5 in 7000
Cuthbertson (1908)	6628	± 7 to ± 21 in 7000
present work (1942)	6627	± 1·7 in 7000

(*c*) *Carbon oxysulphide* (COS)

No previous measurements have been made on this gas, though for the purpose of linking refractivities with structural properties, it is necessary to have refraction and dispersion data available, since already data exist for carbon dioxide (CO_2) and carbon disulphide (CS_2).

Carbon oxysulphide has been prepared on many occasions by previous investigators for the purpose of determining the chemical properties of the gas. Some measurements have also been made on physical properties but there is no doubt that the descriptions of the methods given in the literature leave much to be desired, indeed most workers give little or no particulars as to their methods of production of the gas, being content to refer simply to the text-books. Gonzalez & Moles (1919) state that 'the preparation is always *assez mediocres*', whilst Russell (1900) writes: 'I have not found any method of purification which can be relied upon to give a sufficiently pure gas'

Many weeks were spent in trying out the various methods of preparation previously used, but in nearly all cases it was found that the refractivity determinations proved that the various products were impure, consequently a new method was evolved based upon these experiences with the object of eliminating certain undesirable features of existing methods. It may be pointed out here that there is no adequate method of chemical analysis of the final gas product which can even remotely approach the sensitivity of the interferometer measurements and so no analysis was attempted. The purity of our gas product was estimated from the uniformity of the refraction measurements made on various samples obtained at various times.

The main difficulty in the preparation of carbon oxysulphide arises from the hydrolysis of the gas, since in the presence of the slightest trace of moisture, it changes readily into CO_2 and H_2S. Hence it is essential to effect immediate drying at each stage. The final arrangement adopted for the preparation was as follows, based on the method of Klason (1887): 50 c.c. saturated KCNS solution + 290 c.c. conc. H_2SO_4 + 400 c.c. distilled water (cold), the gas being passed through (*a*) one wash-bottle containing 33 % NaOH solution (to remove CO_2), (*b*) one tube of conc. H_2SO_4 on pumice (for drying), (*c*) four U-tubes of solid NaOH (to remove CO_2), (*d*) one tube of solid mercuric oxide (to remove H_2S), (*e*) three bottles of conc. H_2SO_4, one tube P_2O_5, one tube $CaCl_2$, one tube P_2O_5. The last section (*e*) serves to complete the drying and removal of any final traces of impurity.

The gas was solidified in a U-tube immersed in liquid oxygen, fractionated into a second tube, and finally fractionated into the apparatus. Results obtained in the refraction measurements are as follows:

pressure coefficient	$a = 0{\cdot}1339 \times 10^{-4}$ per mm.
theoretical density	= 2·0749 (air = 1)
experimental density	= 2·1046

Temperature coefficient $b = 0{\cdot}003830$ per °C. No previous measurements of these coefficients appear to have been made.

TABLE 6. DISPERSION OF CARBON OXYSULPHIDE (COS)

λ	$(n-1)_0 \times 10^7$	λ	$(n-1)_0 \times 10^7$
5792 26	8817·0	4801·25	8809·4
5771·20	8819·3	4679·46	8924·1
5462·25	8842·9	4359 56	8984·6
5087·23	8874 0	4047·70	9061·9

The Sellmeier formula corresponding to table 6 is

$$(n-1) = 9{\cdot}372 \times 10^{27}/(10{,}903 \times 10^{27} - \nu^2).$$

The probable error E in the determination for the mercury green line is $\pm 0{\cdot}4$ in 9000.

Our thanks are due to Mr H Holness, M Sc., A.I.C , Senior Lecturer in Chemistry in the College, for his practical assistance with the preparation of the gases, in particular with devising the method for the purification of COS; and to Mr Derek Price, B.Sc., for assistance in all stages of the experiments. We also acknowledge with thanks the loan of certain apparatus from the Government Grant Committee of the Royal Society.

REFERENCES

Cuthbertson & Metcalfe 1908 *Proc. Roy. Soc.* A, **81**, 440.
Cuthbertson, C. & M. 1910*a* *Proc. Roy. Soc.* A, **83**, 151.
Cuthbertson, C. & M. 1910*b* *Proc. Roy Soc.* A, **83**, 171.
Cuthbertson, C. & M. 1932 *Proc. Roy. Soc.* A, **135**, 40.
Fajans, Wust & Reindel 1934 *Z. phys. Chem* **24**, 103.
Gonzalez & Moles 1919 *J. Chim. phys.* **17**, 408.
Ketteler 1885 *Theor. Optik.*
Klason 1887 *J. prakt. Chem.* (2), **36**, 66.
Lowery 1927 *Proc. Phys. Soc.* **40**, 23.
Lowery 1931 *Proc. Roy. Soc.* A, **133**, 188.
Mascart 1874 *C.R. Acad. Sci., Paris*, **78**, 717.
Mascart 1878 *C.R. Acad. Sci., Paris*, **86**, 321.
Meggers & Peters 1918 *Sci. Pap. U.S. Bur. Stand.* no. 327, p. 697.
Sears & Barrell 1939 *Phil. Trans.* A, **231**, 75.
Stuckert 1910 *Diss. Karlsruhe.*
Walker 1903 *Phil. Trans.* A, **201**, 435.

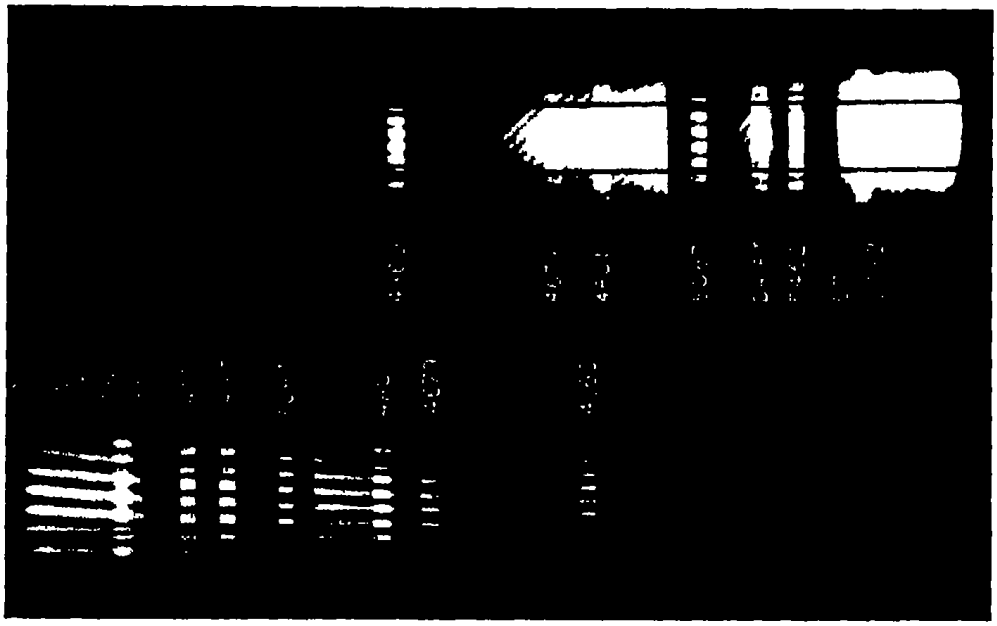


FIGURE 1

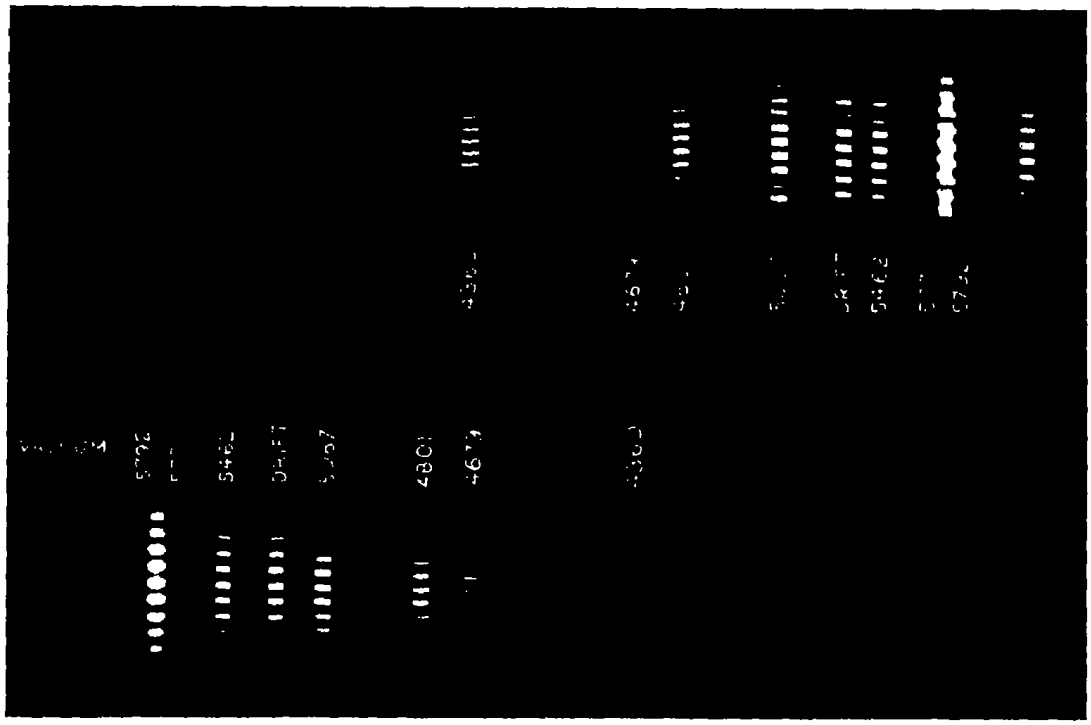


FIGURE 2

(*Facing p* **216**)

Address of the President Sir Henry Dale, G.B.E., at the Anniversary Meeting, 30 November 1943

Before proceeding to the presentation of the Medals awarded for this year, it is fitting that, in accordance with custom, we should briefly recall the lives and the achievements of those whom death has removed from our Fellowship and our Foreign Membership since the last Anniversary Meeting of the Society.

War conditions have delayed till now the receipt in this country of news of the death, two years ago, of HENRI LÉON LEBESGUE (1875–1941) Lebesgue initiated a new concept of the integral in his work on the theory of functions of a real variable, which has had a revolutionary effect in various branches of mathematical analysis. At one time Professor of Mathematics in the University of Poitiers, he spent all his later life as Professor of Mathematics in the University of Paris. He was elected a Foreign Member in 1934

EDGAR JOHNSON ALLEN (1866–1942) will long be remembered for his outstanding services to marine biology. For 42 years he was Director of the Marine Biological Association's Laboratory at Plymouth, and the steady growth of this institution, through years of adversity to its present leading position, is mainly due to his wise guidance and life-long devotion.

Enthusiasm for marine research was the inspiration of Allen's life, and in a number of directions his work has proved to be the starting point from which further active investigation has proceeded It was his aim to make Plymouth a centre of marine research, and to this end all his energies were primarily directed. In spite of financial difficulty he succeeded in providing accommodation, a suitable ship, an unusually fine library and all the other appurtenances which were needed, and year after year, in growing numbers, students from all parts of the world came to work at Plymouth. Allen's exceptional knowledge of every branch of marine biology was freely available to all who needed help or advice, and a spirit of helpfulness and goodwill pervaded the laboratory under his direction. Allen shunned any form of personal publicity, and his whole life was devoted to the encouragement of science. He was elected a Fellow of the Society in 1914 and was awarded the Darwin Medal in 1936.

HENRY MIERS (1858–1942), the senior and, by general consent of his colleagues, the most widely known and versatile of our Fellows in the field of mineralogy and crystallography, was elected to our fellowship in 1896, soon after his induction to the new Waynflete Chair of Mineralogy at Oxford. He had graduated at Oxford fifteen years earlier with honours in mathematics, after entering the University as a classical scholar from Eton.

The majority of Miers's long list of published papers were devoted to crystallography, and the most conspicuous among them was perhaps the memoir which was issued in our *Philosophical Transactions* forty years ago under the title of 'An enquiry into the variation of angles observed in crystals'. By studying crystals during their actual growth, with the help of a novel form of goniometer, Miers showed that the small variations in the recorded interfacial angles of crystals were not mere 'imperfections of nature', but were due to the alternating growth of vicinal faces strictly in accordance with the fundamental law of rational indices, although they differed widely from the commonly recognized axial ratios.

Miers's remarkable flair for administration, rendered easy and acceptable to colleagues by his generous and genial nature, resulted in frequent interruptions of his dominant enthusiasm for teaching and research, the two most serious of these being seven years spent as Principal of the London University and a similar period as Vice-Chancellor at Manchester. The latter part of a happy life of 84 years, devoted unselfishly to the public interest, was occupied with constructive surveys of public museums and libraries in this country, Canada, the United States, continental Europe and South Africa. He was knighted in 1912.

Frank Dawson Adams (1859–1942) was born in Montreal and descended from an old colonial family of New England of which one member was John Quincy Adams, a famous President of the United States of America. Educated at McGill University, Montreal, he early decided to devote himself to the study of geology, mineralogy and petrography, and, after graduating, proceeded to Harvard and thereafter to Heidelberg and Zurich. Subsequently he joined the Staff of the Geological Survey of Canada, on which he served as Assistant Chemist and as Petrographer. On Sir William Dawson's retirement in 1898 Adams succeeded him as Logan Professor of Geology in McGill University, a post he held until his retirement in 1924. Adams specialized in the study of the Archaean rocks of the Canadian Shield, and from the first his papers were marked by ripe scholarship and great originality. In his petrology, even when opposed to current opinions, he made practically no mistakes, and his results have nobly stood the test of time. Availing himself of the resources of the engineering laboratories of McGill, he embarked on a long and very laborious series of experiments to determine the plasticity of rocks such as marble, granite and basalt under the highest stresses available in experiment. Adams was regarded as one of the outstanding men of science in the Dominion and in the western hemisphere and received many honours. He was elected a Fellow in 1907.

Arthur Willey (1867–1942) was one of the last representatives of that extraordinarily fruitful period of zoological research that followed the work of Charles Darwin. The fascinating idea that clues to the ancestral history of an animal group or of an individual organ are to be found lurking in the body of the embryo or larva lighted a flame of enthusiasm for zoological research, and that science became, for the time being, preponderatingly morphological. Willey followed the

enthusiasm of the day, and his research work—much of it of great and permanent value—centred round the morphology of various supposedly 'primitive' types. First came *Amphioxus*, which Kovalevsky's embryological investigations had shown to be undoubtedly allied to the vertebrates, then other relatives, the Ascidians, greatly degenerate as a result of leading a motionless sessile existence; and those strange worms the Enteropneusta, for which Bateson had claimed vertebrate affinity. Later came Willey's great expedition to the East Indies in search of embryos of the pearly *Nautilus*, fruitless as regards its main quest but yielding a rich store of rare and interesting marine animals which provided the material for a long series of important publications. As teacher in Columbia University, as Head of the Colombo Museum, and finally as Professor of Zoology at McGill, Willey everywhere rendered able and faithful service. He was elected to our fellowship in 1902.

Charles Tate Regan (1878–1943), late Director of the British Museum (Natural History) and an ichthyologist of world-wide reputation, entered the British Museum in 1901 as assistant to G. A. Boulenger and, after a short time, was put in sole charge of the Fish Department. During the succeeding thirty-seven years he served the Museum in posts of increasing importance and eventually, in 1911, became Director, retiring in 1938 at the age of 60.

Tate Regan held strong views on museum organization, being a consistent advocate, both before and after he became Director, of the paramount importance of the research side of the Museum's activities. Throughout his working life he acted up to this ideal and himself carried out a large volume of research of high value on the classification of fishes. His first major work (a monograph of the Loricariidae) was published in 1904. Similar revisions of other families of fishes, extensive studies of larger groups, reports of the results of expeditions to various parts of the world, and shorter papers on new genera and species (totalling in all more than 250 separate publications)—these followed one another in unbroken sequence and steadily built up his reputation as one of the leading systematic ichthyologists of the day. He was elected a Fellow in 1917.

By the death of Alexander Russell (1861–1943) the Society loses an applied mathematician who gave almost his whole attention to the mathematical solution of problems arising in electrical engineering. He obtained his mathematical education at the University of Glasgow, and at Caius College, Cambridge. As a young man he joined the staff of Faraday House, a London College established for the training of engineers by thoroughly practical methods. His published scientific papers nearly all deal with what may be called the geometry of the fields of force associated with electrified conductors, or conductors carrying electric currents. He derived a number of much-used formulae for the self- and mutual inductances of electric cables, and for the electric capacity of conductors such as nearly concentric spheres, or spheres nearly or just in contact. He has put on record a number of unexpected results, such as the theorem that, in the

case of a 'concentric' main, there is no mutual inductance coefficient between outer and inner conductors when their axes are not coincident. He was for many years Principal of Faraday House, and was a past President of the Institution of Electrical Engineers and of the Physical Society. He was elected a Fellow in 1924.

CLINTON COLERIDGE FARR (1866–1943), born at Adelaide, made valuable contributions to many branches of physics, among which may be mentioned high-tension insulation, electric charges in the open air, seismograms, the porosity of porcelain and the viscosity of sulphur. But he will chiefly be remembered for the complete magnetic survey of New Zealand and the outlying islands and for the magnetic observatory established by him at Christchurch, N.Z His first post, that of Lecturer in the University of Sydney (1891–1895), was followed by an appointment to lecture on electrical engineering in the University of Adelaide. From 1898 to 1904 he was engaged on the magnetic survey of New Zealand. The remainder of his active life was devoted to the University of New Zealand, where he was appointed Lecturer (1904) and Professor of Physics (1910–36) in Canterbury University College, Christchurch He was elected a Fellow in 1928.

FRANK STURDY SINNATT (1880–1943) was Director of Fuel Research to the Department of Scientific and Industrial Research. To his perseverance and enthusiasm was largely due the growth of the National Coal Survey, until it included 98 % of the coal in the United Kingdom. He had a very wide knowledge of the characteristics of British coals, and his efforts made an important contribution to their proper utilization As Lecturer in Applied Chemistry at the Manchester College of Technology he made many investigations on coal, particularly on the formation of cenospheres. He helped to found the Lancashire and Cheshire Coal Research Association, and to develop the methods used in the coal survey. As Superintendent of the Survey and Director of Fuel Research he was responsible for many investigations on the breaking of coal, coal cleaning, carbonization, hydrogenation, etc. He was a distinguished servant whom the state could ill afford to spare. He was elected a Fellow in 1938.

SIDNEY GERALD BURRARD (1860–1943) entered the Royal Engineers in 1879. He went to India in 1882 and in 1884 joined the Survey Department He was posted to the Trigonometrical Branch in which he was to have so distinguished a career. For the first years of his service he was engaged on astronomical determinations of latitude and longitude. In the years 1894–6 he was one of the two observers engaged on the redetermination of the longitude of India, starting from the Prime Meridian at Greenwich and carrying the work across Germany, Russia, Persia and the Persian Gulf. His greatest work, namely, the investigation of the causes of deflexions of the plumb-line, had been going on for several years and was given to the world in 1901 when he published the paper called 'The attraction of the Himalaya Mountains on the plumb-line in India'. This paper changed the ideas that had been held for half a century and led to a

great renewal of interest in geodetic research, showing, as it did, that much light on the structure of the earth's crust could be derived from a study of plumb-line deflexions. In 1911 Burrard was appointed Surveyor-General of India and became a worthy successor of Everest and J. T. Walker. In 1914 he was created K.C.S.I.

Burrard came of a family of Norman origin which had concentrated near Lymington about the end of the fourteenth century and had remained there for five centuries. In 1769 the then head of the family, who had been M P. for Lymington for 44 years, was made a Baronet In 1933 this title was inherited by Sir Sidney from a cousin, and he thus became the seventh Baronet.

Since scientific research became a specialized occupation, it has been less common for amateurs of science to be elected as Ordinary Fellows of the Society. Edward Heron-Allen (1861–1943) belonged to the legal profession by birth and training, and his firm practised for a number of years as Parliamentary Agents on behalf of the City of Westminster. Throughout his life, however, most of his interests and activities were concerned with other subjects than the law In his leisure time he became a skilled violin maker and wrote a treatise on the history of violin construction Another of his recreative interests was Persian literature As a naturalist, Heron-Allen concentrated attention for many years upon the collection and study of fossil and living Foraminifera. As a result of his own experience he published, fifty years ago, a paper on the cretaceous Foraminifera which became the leading practical guide on the technique of preparation, in the study of fossil forms of the Order. From 1907 to 1933 he was associated with Mr Arthur Earland in the collection and description of many species of these simple organisms, the range of which in time and space is very remarkable.

In 1925, Heron-Allen presented to the British Museum (Natural History) the whole of his collection of Foraminifera, in which many other notable collections had been included, and his extensive library relating to the order. He accepted the post of Honorary Curator at the Museum He was elected a Fellow in 1919.

Edwin John Butler (1874–1943) studied at Queen's College, Cork, in medicine and in botany After holding a travelling scholarship in botany, he served in India for nineteen years, first as Cryptogamic Botanist to the Government (1901), than as Imperial Mycologist (1905), Joint Director of the Agricultural Experiment Station, Pusa (1919) and Agricultural Adviser to the Government (1920). During his Indian career he established a great reputation both in original research and in administration, and, owing to the immense area covered, from the temperate north-west to the tropical south, he gained an unrivalled experience of the plant diseases of crops grown under the most various conditions.

In 1920 he returned home to organize the Imperial Bureau of Mycology (now the Imperial Mycological Institute), and for fifteen years was its Director, and the Editor of the *Review of Applied Mycology*. During this time he largely influenced the co-ordination of plant pathology especially in the tropical parts of

the Empire, and the *Review* became indispensable to phytopathological stations round the world.

In 1931 he was appointed one of the first members of the Agricultural Research Council, and in 1935 became the Secretary to the Council, serving until 1941. Among his many activities may be mentioned his part in the organization of the grain infestation survey and his service on the Technical Committees on Plant Virus Diseases and on Fungicides and Insecticides. He became a Fellow of the Royal Society in 1926, and was knighted in 1939.

WARRINGTON YORKE (1883–1943) was a graduate of Medicine of Liverpool University, where for a short time he was Holt Fellow in Physiology under Sir Charles Sherrington. In 1907 he joined the Liverpool School of Tropical Medicine, which he served for the rest of his life, first as Director of the Runcorn laboratories, then as Walter Meyer Professor of Parasitology in succession to Ronald Ross, and finally as Alfred Jones Professor of Tropical Medicine in succession to J. W. W. Stephens Yorke served on expeditions to Africa to study blackwater fever and trypanosomiasis. The work in Rhodesia proved that the tsetse fly, *Glossina morsitans*, was the transmitter of the Rhodesian sleeping sickness, and that wild mammals formed a reservoir of the infection, almost simultaneously with the similar discoveries of David Bruce with the Royal Society's Commission in Nyasaland. On the purely parasitological side Yorke made many valuable contributions to knowledge, perhaps the most important of which was his text-book on the nematode parasites of vertebrates, written in collaboration with Dr P. A. Maplestone.

During the last ten years of his life he turned his attention more especially to chemotherapy. His researches threw light on the action of specific drugs and the phenomenon of drug fastness, and some of the most recent had opened up a completely new field of chemotherapeutic agents, the diamidines, some of which have proved to be remedies of great value for kala azar, one of the most fatal and widespread of tropical diseases In the midst of this work, and knowing that he was stricken with a fatal malady, Yorke had begun to prepare a Croonian lecture which he was to have delivered to the Society, and accepted an arduous mission to the U.S.A., to co-ordinate research and other activities in regard to remedies of urgent importance for the war. His death has deprived the Liverpool School and British research in tropical medicine of a progressive and inspiring leader. He was elected a Fellow in 1932.

DAVID HILBERT (1862–1943) had been a Foreign Member since 1928, and was by general admission one of the greatest mathematicans of our time. He was professor at Göttingen from 1895 to 1932, and it was primarily to him that Göttingen owed the position which it occupied for many years as the first international centre of mathematics in the world.

There is hardly a branch of pure mathematics to which Hilbert did not make contributions of the first importance, and there are several which he transformed

entirely. His greatest constructive work, in algebra, number theory, the foundations of geometry, the calculus of variations, and the theory of integral equations, was done before 1914. Since then he had been primarily a logician, and it may be too early to pass a final judgement on his work in that field, but there is no doubt that a great deal of the logical activity of the last twenty years is the result of his ideas and inspiration.

THOMAS HUDSON MIDDLETON (1863–1943) was one of the band of pioneers whom agricultural science was fortunate enough to attract at the time when public funds were beginning to be made available for its development. He came of a long line of good farmers in the north of Scotland and was deeply imbued with the spirit of the old agricultural improvers. He began his university career at Glasgow, where he studied engineering; then he went to Edinburgh to study agriculture. Later he had a distinguished career as Professor of Agriculture, first at Baroda, then successively at Aberystwyth, Newcastle, and Cambridge. He then joined the Ministry of Agriculture as Scientific Adviser, but in 1917, when the food situation in this country had become critical, he was put in charge of the technical side of the Food Production Department. Here he originated and was responsible for the ploughing-up policy which gave such admirable results then and now. Afterwards he joined the Development Commission and later became Chairman of the Agricultural Research Council. He played a great part in the development of agricultural education, research, and advisory activities in this country. He was knighted in 1929 and was elected a Fellow in 1936.

KARL LANDSTEINER (1868–1943), Member of the Rockefeller Institute for Medical Research, was born in Vienna and migrated to the United States in 1922. After taking a degree in medicine, he studied organic chemistry under Emil Fischer before beginning his life's work in bacteriology and immunology. This training had an influence on the methods and angle of his approach to the problems of immunity and enabled him to play a leading part in the development which has brought these problems into relation with more precise conceptions of chemistry and physics. In the field of investigation dealing with virus infections he was also a pioneer, in demonstrating the transmissibility of poliomyelitis to monkeys. His discovery of group differences in human blood was recognized by the award of a Nobel Prize in 1930. His book on *The Specificity of Serological Reactions* not only records the results of a long series of brilliant researches, but places them in their proper perspective in a wide field of biological and physico-chemical ideas. He was elected a Foreign Member in 1941.

WILLIAM ROBERT BOUSFIELD (1854–1943), barrister and man of science, was born at Bedford, where he attended the Modern School. He served an apprenticeship as an engineer, and then went to Caius College, Cambridge, where he won scholarships and took his degree as sixteenth wrangler in 1876. After lecturing on mathematics and engineering at Bristol, he was called to the Bar in 1880,

and soon became known as a specialist in patent law. From 1892 to 1905 he was a Conservative Member of Parliament, but about 1902 he became interested in research and began a series of investigations, mainly on the theory of solution, partly in conjunction with the late T. M. Lowry.

Bousfield's researches dealt with osmotic pressure and allied phenomena, ionization and the law of mass action, and especially with the sizes of ions. These sizes he estimated by assuming the ions to be spheres, and applying Stokes's law for the motion of a sphere in a viscous fluid. In order to account for the experimental data for binary electrolytes, it was necessary to suppose that ions are molecular aggregates composed largely of water, the sizes changing when temperature or concentration is altered. He was elected a Fellow in 1916

Gerard Jakob de Geer (1858–1943) was elected a Foreign Member in 1930. His great achievement has been the deciphering of a year-chronology backwards into a part of the earlier stone age. The first idea, which came to him in 1878, was based on the recognition of annual layers of growth in the 'varved' clays of the Baltic, in which the laminae may be compared with the annual rings recognizable in trees of temperate regions. From then onwards to the end of his life, with the help of assiduous collaborators, every difficulty was overcome, every relevant line of inquiry brought to aid, and the work extended to North America and other parts of the world. He was Professor of Geology in the University of Stockholm from 1897 to 1924, and then became Director of the Geochronological Institute of the University, established in that year for the new subscience which he had created, and which has added such unexpected precision to glacial geology.

Henry Guy Ellcock Pilgrim (1875–1943) was born at Stepney, Barbados, and died at Upton, near Didcot, Berkshire. He was educated at Harrison College, Barbados, and University College, London (D.Sc. 1908). He was appointed to the Geological Survey of India in 1902, and retired with the rank of Superintendent in 1930. He then became a member of the Supernumerary Staff of the British Museum (Natural History).

Pilgrim's work can be divided into two main sections, as a geologist and as a palaeontologist As a geologist he worked in the field in Persia and Arabia, and in the Tertiary regions of northern India. In the Persian Gulf region he established the sequence of formations from pre-Cambrian to Recent, whilst in India his researches on the fauna of the Siwaliks enabled him to subdivide this formation on the basis of its fossil vertebrates, and to lay down a classification, which, like that of the formations of the Persian Gulf, has been found trustworthy by later workers.

As a result of his studies of the vast vertebrate fossil faunas of the Siwaliks, Pilgrim became one of the most distinguished exponents of vertebrate palaeontology. His results are contained in a magnificent series of Memoirs published in the *Palaeontologia Indica*, and in Catalogues of the British Museum. Special mention may be made of his discovery in the Siwaliks of five new species and two new

genera of fossil primates, of which *Sivapithecus* was regarded by Pilgrim (1927) as the anthropoid closest phylogenetically to man, with the possible exception of *Pithecanthropus* and *Eoanthropus*. He was elected a Fellow this year, but had not been admitted.

Pieter Zeeman (1865–1943) was born at Zonnemaire, a village in Zeeland. He studied physics under Kamerlingh Onnes and H. A Lorentz at Leyden, where he afterwards became assistant in the Physics Institute. In 1897 he went to Amsterdam, where he remained for the rest of his life, he was attached to the University first as lecturer and three years later as Professor of Physics. The well-known effect which bears his name brought him the award, jointly with H A. Lorentz, of a Nobel Prize in 1902 for 'researches on the influence of magnetism on radiation phenomena'. The new discovery was published in papers on the influence of magnetism on the nature of light emitted by a substance (1896) and on doublets and triplets in the spectrum produced by external magnetic forces (1897). He wrote many other papers on magneto-optics, developing the subject in detail He was elected a Foreign Member in 1921, and awarded the Rumford Medal in 1922

Geoffrey Thomas Bennett (1868–1943), for just fifty years a Fellow of Emmanuel College, Cambridge, has contributed to a wide range of mathematical subjects, ranging from the theory of numbers to the geometry of mechanism, which latter might be regarded as his special subject. Many of his contemporaries remember the help which he readily gave to mathematical problems arising in many different branches of science. During the 1914–18 war he worked on naval and anti-aircraft gunnery at Whale Island, and later at the Admiralty compass observatory at Slough He made an extensive collection of mathematical models at Emmanuel College. He was elected a Fellow in 1914.

William Herbert Hatfield (1882–1943) was a Sheffield man, and one of the earliest graduates in Metallurgy of the University of Sheffield, under Professor J. O. Arnold. His first industrial experience was in the steel works of Sir Henry Bessemer and Co., but later he went to Messrs Crowley's foundries, of which he became a director. Here he published his first original research, a study of the open-hearth process, jointly with Professor A. McWilliam. In 1916 he went to the Brown-Firth Research Laboratories, of which he was the active head until his unexpected death During this time he built up an important research organization for his firm, and his name is particularly associated with the development of the austenitic stainless steels containing nickel and chromium, the properties of which he studied with great thoroughness. The Iron and Steel Industrial Research Council grew from a group of research committees which, under Hatfield's chairmanship, sponsored a large and important body of researches on ingots, alloy steels, corrosion, and allied subjects. The success of these investigations, recorded in a long series of published papers, was largely due to the stimulus given by

the chairman. During the war, as chairman of the Technical Advisory Committee on Special Alloy Steels, he was instrumental in bringing about a reduction of the large number of specifications for such steels to a selected short list, and thus saving quantities of valuable alloys and simplifying the task of the producers of armaments. The strain of these additional efforts, and of a recent visit to America as one of a metallurgical mission, probably told on his health, so that he died early. He was elected a Fellow in 1935.

Clive Cuthbertson (1863–1943) was born in Shanghai and passed his early years in China, Burma and Australia with his father, a merchant and banker. In 1878 he went to Glenalmond and passed into the Indian Civil Service, proceeding to Bengal in 1885. Here he developed tuberculosis and he had to retire in 1896. After two years as an invalid he became interested in science and attended a course at University College, London, with the intention of taking up research. He had observed a singular relation in the values of the refractivities of the inert gases. Following this up, he completed the work on elementary gases and vapours and established an important relation connecting the refractivities of the elements. Most of his work was published in the *Proceedings* and *Transactions* of the Society.

He was at one time (1900–1902) private secretary to the prime minister, Lord Salisbury. From 1915 to 1919 he worked at the Foreign Office, work recognized by the award of an O.B.E. He also published a book on the currency question at the time of the bimetallism controversy. He was elected a Fellow of the Royal Society in 1914.

Edward Bagnall Poulton (1856–1943) studied zoology at Oxford under Rolleston; and under Moseley's influence took up morphological studies which resulted in the publication in 1888 of the discovery of true teeth in *Ornithorhynchus*. Attracted by Weismann's essays on heredity, he took up the study of the coloration of insects, and his book, *The Colours of Animals* (1890), contained the first comprehensive classification of all forms of coloration, with logical nomenclature for them. It brought him to the front rank of evolutionists. Throughout his life he vehemently supported the theory of natural selection, opposing the claims of saltation of the early mutationists. As Hope Professor of Zoology at Oxford, 1893–1933, he got together a band of naturalists, through whose field work, mainly in tropical Africa, an immense body of knowledge was published in notes and studies on all forms of coloration in insects, the prey of predators, courtship, polymorphism in butterflies and their geographical races.

He was elected Fellow of the Royal Society in 1889 and received the Darwin Medal in 1914.

He presided over the second International Entomological Congress at Oxford in 1912, the Linnean Society 1912–1916, received the Linnean medal in 1922, and was thrice president of the Entomological Society, the last occasion being at the centenary meeting in 1933, when it became a Royal Society. He was knighted in 1935.

Awards of Medals, 1943

The COPLEY MEDAL is awarded to Sir JOSEPH BARCROFT, C B E., Emeritus Professor of Physiology in the University of Cambridge, and now the head of a research unit established there in Animal Physiology.

The researches on respiration and the respiratory function of the blood, in respect of which this highest of the distinctions in the Society's gift is awarded this year, have occupied a central position in Barcroft's life-work. Their record is one of steady purpose and unbroken progress over a period of some 45 years already. At Barcroft's own hands, and at those of the long and very distinguished succession of collaborators who have been associated with him, these researches have had many important extensions into neighbouring fields of biological enquiry; they have played a great part in a wide range of recent advances in physiological, biochemical and medical science throughout the world, and they have been an important factor in the creation of a scientific basis for the study of manifold human activities, of work and of play, in peace and in war.

As a newly graduated student at Cambridge, Barcroft seems to have seen already the significance of the problem to which he was thus to devote so many years of fruitful investigation, recognizing the respiratory function of the blood as a condition of the life and a measure of the metabolic activity of every part of the body of the higher animals. With a cumbrous equipment of pumps and their accessories, alone available for such determinations in those early days, he applied himself with great determination and ingenuity to a study of the changes of oxidative metabolism which accompany secretory activity in the salivary gland. From that beginning his researches grew and developed in logical sequence, including studies of oxidative metabolism in relation to specific activity in the kidneys, the liver, the muscles and other organs, with various excursions to deal with corollaries from this central theme

Early in the series came Barcroft's invention of a comparative manometric method, which effected a revolutionary simplification of the technical operations involved in obtaining the data required The same method enabled him, in due course, to make his classical redeterminations of the relation between the pressure of oxygen and its combination with or release from blood or solutions of haemoglobin, and of the effects of varying conditions on the graphs representing that relation. The principle of this method, as employed either in Barcroft's own original apparatus, or in the form which Warburg later introduced, has passed into standard use in every laboratory of functional biology throughout the world, for measuring rates of respiratory metabolism in surviving samples of living tissues and of various chemical changes due to artificial enzyme systems. Barcroft and his companions in research have used it, in expeditions which he led to Monte Rosa and to the Andes, to study the modifications of man's physiology by life at high altitudes and at low oxygen tensions, providing knowledge which to-day has

an ever more urgent application to the problems of human flight, and in two successive wars Barcroft himself has used the knowledge built up and the skill acquired in his peaceful researches to the direct aid of his country in its need.

Studies of the differences between the respiratory functions of the blood in foetal and independent life led Barcroft, with a characteristic readiness to follow the lead of natural interest into a neighbouring field of enquiry, to a more general study of the physiology of the mammalian foetus, and thus to the results which he reviewed here in the Croonian Lecture of 1935 on 'Foetal respiration' His need, for that purpose, to make observations on larger animal types, such as the sheep, led again to a wider interest in the physiology of our farm animals, and so to the new enterprise for which his more recent research appointment provides welcome opportunity.

In a preface to his book, Joseph Barcroft finds a connexion between what he learned in sailing boats and what he has done in research. We may think of him, perhaps, as one whose instinct and aptitude have led him to venture as an explorer beyond visible horizons, rather than to sound remoter depths of theory in seas already known Such a thought would place him in a long tradition of this Society, in company with such great students of the respiratory function as Boyle, Hooke, Mayow and Lower among our earliest Fellows, and, of those who over the centuries have received this Copley Medal, with Stephen Hales and Joseph Priestley and, most intimately of all, with John Scott Haldane who received it nine years ago. Barcroft's work, like Haldane's, of high scientific merit in itself, has extended its influence far beyond the laboratory over a wide range of the beneficent and creative activities of mankind at peace, and now again, for the last time let us hope, to those imposed by the dire compulsion of war. We rejoice to know that in him the hunger for knowledge is still unsatisfied, and that he is still sailing his craft towards new horizons.

A Royal Medal is awarded to Sir Harold Spencer Jones, Astronomer Royal for England, in recognition of his determination of the solar parallax and of other fundamental astronomical constants

Much of the earlier work of Spencer Jones was devoted to the discussion of certain fundamental astronomical constants, and in this field of work he had long been recognized as a leading authority. At the meeting of the International Union held at Leyden in 1928, a Solar Parallax Commission was appointed, with Spencer Jones as President, in preparation for the favourable opposition of the minor planet Eros in 1931.

The planning and coordination of the enterprise of redetermining the solar parallax, in which 24 observatories took part, fell mainly upon him, and he was entirely responsible for the assemblage of the material, and for the derivation of the results. It is to be noted also that the Cape Observatory, of which he was then director, made the greatest contribution to the observational material.

The circumstances of the opposition favoured southern observatories, and the plates obtained at the Cape contributed more weight than those from all the other observatories combined

The whole task of reduction of the observations was carried out under Spencer Jones's supervision, and the memoir which contains the complete discussion of the material and the results obtained from it (*Mem. R. Astron. Soc.* 1941, vol. 66, part 2) is a masterly exposition of the whole subject of the investigation of solar parallax and of the inter-related constants

The solar parallax finally derived is given as $8{\cdot}790 \pm 0{\cdot}001$, as compared with the value of $8{\cdot}806 \pm 0{\cdot}004$ obtained by A. R Hinks at the preceding opposition of Eros in 1900–1 Spencer Jones found complete agreement between the results derived separately from northern and southern observatories, or from right ascensions and declinations separately, or from photographic and photovisual telescopes, or from primary and secondary stars, as well as from four successive periods of time during the opposition No matter how the material was subdivided, no substantial difference could be found in the value of the solar parallax derived. There is thus every reason to think that systematic error has been eliminated, and that the probable error of 0·001 truly represents the high degree of accuracy of the final result Therewith we are given a new estimate of the earth's distance from the sun, which is likely to stand without revision for many years. These observations on Eros also yielded a new estimate of the mass of the moon, which has less than half the probable error of the previous determination

The modification of the previously accepted value of the solar parallax affects the estimates of other astronomical and geophysical constants. Spencer Jones's work is enhanced by his discussion and general adjustment of the values of all these inter-related constants, and in this he makes a substantial new contribution to the problem to which so much of his earlier work was devoted.

A Royal Medal is awarded to Dr E B. Bailey, M.C., the present Director of the Geological Survey of Great Britain, in recognition of his contributions to the knowledge of mountain structure and his studies on the tectonics of vulcanism.

The number, range, and importance of Bailey's contributions to geology can have few parallels in the history of this science in Britain. It is only necessary to review his major publications to recognize the solid achievements that stand to his credit in the diverse fields of petrography, tectonics, stratigraphy, glaciology and physiography But it is more particularly by his outstanding contributions to the knowledge of mountain structure and on the tectonics of vulcanism that he is to-day acknowledged internationally as a leader.

His work on mountain structure is contained in a long series of memoirs on the non-fossiliferous schists of the Southern Highlands of Scotland. These studies, prosecuted far across country and presented in a brilliant synthesis of the structure

over a large part of the south-west and central Grampians, mark a great advance in the field of mountain tectonics and have provided a notable stimulus to research on Highland geology.

Bailey's now classic researches undertaken in conjunction with Clough and Maufe on the Devonian volcanic cauldron subsidence of Glen Coe, and his early work on the petrology and classification of the Carboniferous volcanic rocks of the Midland valley of Scotland, were stepping stones to the immense task which he set himself in later years in the island of Mull. It is these investigations which have revolutionized our knowledge of the inner structure of British volcanoes; for to them we owe the recognition of ring dykes and ring complexes and the conception of igneous intrusion by ring fracture stoping.

Bailey's insatiable zest for scientific research has led him into other fields, and among notable contributions to Scottish geology must be reckoned the work he carried out with Kendall on the glaciation of East Lothian and the additions he made to the knowledge of the physical geography of Millstone Grit, Kimmeridgian and Cretaceous times. Farther afield his studies have dwelt on the tectonics of the Pennsylvania Piedmont and on the seismic origin of submarine canyons of the continental slope: in relation thereto he has shown the importance to be ascribed to submarine landslips in the past record of the rocks, notably in the Palaeozoic succession of Quebec

Throughout a long and active career Bailey by his enthusiasm, his powers of observation and his interpretative insight has rendered splendid service to the science of Geology and his record of achievement is unsurpassed among living British geologists.

The DAVY MEDAL is awarded to Professor I. M. HEILBRON, D.S.O., in recognition of his many notable contributions to organic chemistry, especially to the chemistry of natural products of physiological importance.

After some early researches dealing with the chemistry of the coumarins, chromanes and benzopyrylium salts, Heilbron made his first really important contribution by his well-known investigation on the remarkable hydrocarbon, squalene, which occurs in certain fish oils Almost simultaneously he established the structures of batyl and selachyl alcohols. His interest in fish oils being thus aroused he next turned his attention to vitamin A. He attacked this problem in the first instance by the use of the spectrograph with the collaboration of Morton, who was thereby led to develop methods for the quantitative estimation of this vitamin which have since proved invaluable. Heilbron's contribution to the chemistry of vitamin A earned him an international reputation, and later led to the recognition of vitamin A2. He has attacked also the difficult problem of the synthesis of vitamin A and, although it has not yet reached its objective, his work in this direction has raised many points of great interest and importance. Heilbron has also made notable contributions to our knowledge of the sterols and triter-

penes. In this field, following pioneering work on the structure of ergosterol, he provided an elegant proof of the structure of vitamin D2 (calciferol) and has elucidated the structures of lumisterol, fucosterol and, more recently, zymosterol. In the triterpene field his most notable contribution has been his discovery of the key alcohol, basseol, which, in view of its facile conversion into β-amyrin, seems likely to play a significant role in the classification of triterpene chemistry.

The outbreak of war interrupted Heilbron's projected study of the lipochromes and sterols of the algal classes, a study likely to be of importance not only to the chemist but also to the botanist, for whom it may provide a new basis of classification A résumé of this work formed the subject of a Hugo Muller lecture to the Chemical Society. British chemists are indebted to Heilbron more than to any other chemist for the introduction into this country of the most refined new techniques. He has used extensively in his researches chromatographic absorption, microanalysis and high vacuum distillation, whilst he was probably the first to recognize the potentialities of absorption spectra as a research tool. His influence has thus spread far beyond the circle of the many research chemists trained in his laboratories.

The SYLVESTER MEDAL is awarded to Professor J. E LITTLEWOOD in recognition of his mathematical discoveries.

Among mathematicians the name of Littlewood will always be associated with that of G. H. Hardy for the wonderful example of perfect teamwork in their fashioning of a new and rare instrument of mathematical precision during the second and third decades of the twentieth century. Into the theory of numbers, that subject where problems are so easy to state but so hard to prove, an entirely new spirit was infused when the methods of analysis were brought to bear upon it The credit for attaining the highest results, for perfecting the methods, and for inspiring an enthusiastic following among so many of the ablest mathematicians at home and abroad, all belongs to Hardy and to Littlewood. Littlewood, who is by age the junior in this remarkable partnership, brought his own magnificent contribution as Hardy himself has wholeheartedly testified. Littlewood, on Hardy's own estimate, is the finest mathematician he has ever known. He was the man most likely to storm and smash a really deep and formidable problem: there was no one else who could command such a combination of insight, technique, and power.

Littlewood's work covers a wide field of pure mathematics, extending from his fundamental discoveries in the realm of integral functions (1908) up to the present date. It was a happy event when Ramanujan, with his intuitive genius, proposed problems that brought out the fullest powers and skill of Littlewood and Hardy and their analytical theories. We can have no doubt that there is a peculiar fitness in the immediate succession of Littlewood to Hardy in this triennial award of our Sylvester Medal for mathematical research.

The HUGHES MEDAL is awarded to Professor M. L. E. OLIPHANT in recognition of his work in nuclear physics and mastery of modern methods of generating and applying high potentials.

Oliphant, initially under the general guidance of Rutherford, has made fundamental advances in the field of nuclear disintegration. With the aid of specially designed apparatus he has produced intense beams of protons and of the nuclei of heavy hydrogen, with homogeneous energy, the study of the reactions of which with light nuclei has proved very fruitful. The intensity of the beam, which Oliphant has obtained, has allowed effects to be produced at much lower voltages than had hitherto been available. The theoretically important disruption of the lithium nucleus has been investigated with particular care. In this connexion a separation of the two lithium isotopes was first effected, so that they could be separately studied, with a consequent greater certainty as to the reactions.

In Oliphant's researches the energy liberated in certain nuclear reactions has been measured to a remarkably high degree of accuracy and the laws of conservation of mass-energy and of momentum have been verified. This is a matter of great importance for the understanding of nuclear reactions, and has also enabled the masses of the light atoms to be calculated very precisely.

As a result of work carried out by Oliphant with Harteck, on the bombardment of the nuclei of heavy hydrogen with like nuclei, two new light isotopes have been discovered, one of hydrogen and one of helium, both of mass number 3. The nuclear reactions of beryllium and of boron have also been carefully investigated.

In the published accounts of the work Oliphant's name has often been associated with that of others, including Rutherford, but there is no doubt that he was responsible for the very specialized experimental technique, as well as for other contributions to the collaborative enterprise.

Oliphant has been particularly successful in the design of apparatus. Besides that required for the researches to which reference has been made, he has been responsible for the development of the two-million volt installation at Cambridge During the war he has successfully carried out work of great importance. His fine record of achievement is fittingly recognized by the award of the Hughes Medal for 'original discovery in the Physical Sciences, particularly electricity and magnetism or their applications'.

The Council's Report, covering the period of a year ending 30 September, makes mention of the scientific mission to Australia now completed by our Foreign Secretary, Sir Henry Tizard, whom we are glad to welcome on his return. The Report does not extend, however, to the later departure for India of our Secretary, Professor A. V. Hill The Government of India, through the Secretary of State, asked the Royal Society to depute a distinguished scientist to visit India for con-

sultation on scientific matters, and in particular to advise on scientific and industrial research in relation to measures of post-war reconstruction, and on the co-ordination of such plans in India with corresponding activities here and elsewhere. We felt that our proper response to such an invitation was to let India have a man of the highest qualification from our own Fellowship, and I feel confident that the Fellows will approve of our action in releasing for the necessary period our Senior Secretary, who is also one of our Research Professors, to enable him to accept this important mission. I ask you, further, to send from this meeting a message to Professor Hill of good wishes for the full success of his undertaking, and of hope that one of its results will be to strengthen the bonds of understanding and true comradeship between our Indian colleagues and the men of science of this country. In that connexion I ought further to report to you a step which I have taken, with the approval of the Council, and for which I have not found any precedent in our records. It was brought to my notice that, of the six distinguished Indian men of science who are at present on the roll of our Fellows, only two have hitherto been able to present themselves here in order to subscribe the obligation in our Charter Book and to be admitted according to the statute. It seems certain that the war will create still further difficulty and delay for the attendance here of the other four, and I have accordingly commissioned Professor Hill to take with him to India a sheet of suitable parchment on which the Fellows' obligation is inscribed, and on which signatures can be taken for eventual incorporation in the appropriate page of the Charter Book, unless opportunity should earlier present itself for our colleagues to visit us here and sign directly in the book itself. I have nominated Professor Hill as a Vice-President and, under Statute 42, have deputed him to perform on my behalf our simple ceremony of admission. We hope that he may be able to do this at the meeting of the Indian Science Congress. It seems fitting to take this unusual opportunity thus to complete the reception into the circle of our Fellowship of all the Indian men of science whom the Society has elected. During Professor Hill's absence the Council have invited Dr Salisbury to act as Biological Secretary, and we are grateful to him for consenting to give us this help in the emergency.

Last year we devoted this Anniversary Meeting to a simple celebration such as the war conditions allowed of the three hundredth anniversary of the birth of Isaac Newton. We have noted with appreciative interest that other countries also marked the tercentenary year by paying homage to our Newton's memory. Particular mention is due to the commemorative meetings held, under the tremendous stress of war, not only by the Moscow Academy of Sciences, but also in a number of other scientific centres of Soviet Russia, one as far away as Novo-Sibirsk. The Council's Report mentions the gift which we have sent to the Soviet Academy of Sciences of Moscow, in recognition of this union with our colleagues of Soviet Russia in commemorating one of the greatest scientific achievements of all time, as in the present devotion of all that science can give, in both our countries, to the winning of this war for freedom.

Not since 1941 have I addressed the Society from this Chair, at its Anniversary Meeting, according to regular custom. Only a week after we met here in 1941, the United States of America had become our ally; less than a year later came a turning-point of the war, with Stalingrad and El Alamein; and now the end seems no longer to be in doubt, though we cannot tell how long it may be in coming. I think that it is proper now to claim for science its due share in the achievements which have created the present prospect, in such vivid contrast with that of two years ago. Science in the countries of our great alliance has been devoted without reserve during these two years to the winning of the war, in this country it had been so already for the two years which preceded them. No longer are the allies straining now to overtake a lead gained by the enemy in years of stealthy preparation; the lead is rather on our side. If such things could be weighed and measured, I believe that we should find the alliance to be as far ahead of our enemy in the present volume of our united war researches, and in the brain power of the highest class now concentrated upon them, as in the more readily ponderable output of war material by our industries. And no more than our armed forces or our factories can science afford to relax or to divide its effort until the total victory has been won, without which we can have no faith in the world's future. The increasing certainty of the end, however, imposes upon us with a growing urgency the duty of looking also to that future, and to the part which science must play in the nation and the world when peace returns.

From different influential quarters, as from the Parliamentary and Scientific Committee and from the Federation of British Industries, we have had important pronouncements on the urgent need for national enterprise and national spending on higher education in science and technology, and on the encouragement of research in the applications of science to industry. No body of scientific men will need arguments to convince them that we must think in such matters on a scale, not merely larger, but of a higher order than any with which we have hitherto been familiar. Before the last war Germany had led the world in such development, and between the wars we saw the United States of America move swiftly into the lead. Soviet Russia, starting with a background almost bare of such organization, and with a population largely illiterate, but with leaders having a clear vision of what science was to mean for the modern world, has now shown us what a miracle of scientific education and technical development can be wrought in a quarter of a century. Can it be doubted that another great ally, China, when freed from her long agony of war, will rapidly establish her claim also to high rank among the great nations of the world now in the making? Surely it is clear that, if we are to hold our proper place alongside such great new civilizations, built right from their foundations on modern science, we must ourselves face the problem of giving to science its proper place in the fabric of our own, without grudging or hesitation.

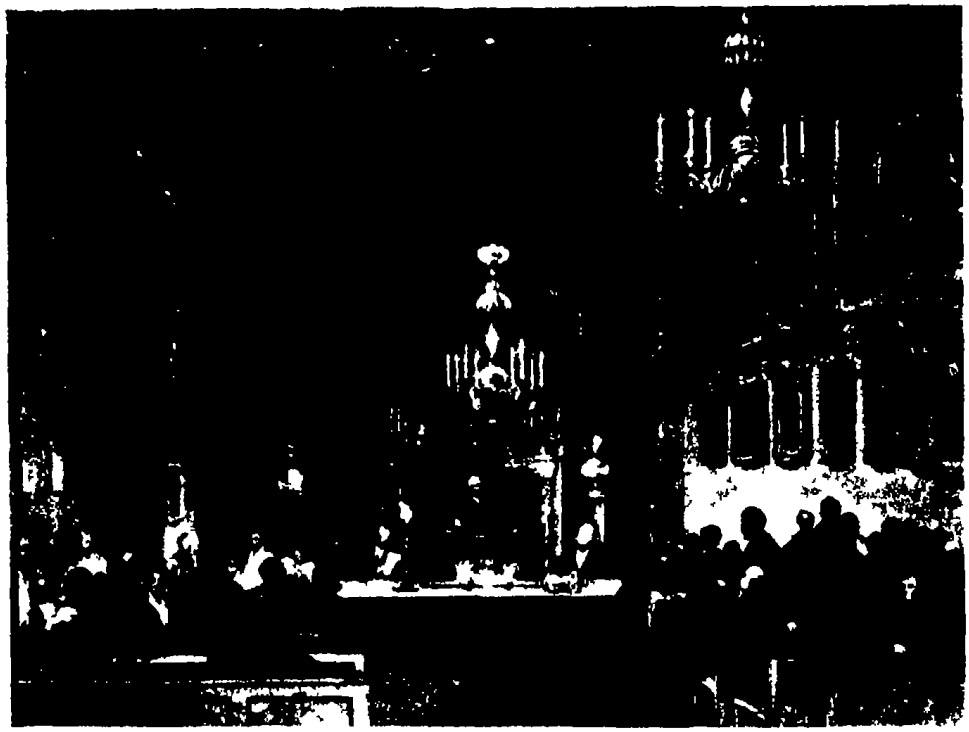
We of the Royal Society shall certainly give enthusiastic endorsement to any movement in that direction. From its beginning to the present day our Society has

always taken a lively interest in the applications of science to the general enrichment of human life, and the enlargement of the means of human happiness. One of the expressed objects of the extension, in recent years, of the number annually elected into our Fellowship, was the maintenance of that interest, under the growing pressure for recognition of achievements in the more fundamental and academic ranges of science. We shall certainly welcome, then, and join in advocating a great expansion of the nation's support of applied science, whether through the Government's Research Councils concerned with researches bearing on industry, medicine and agriculture, or through departments concerned with the uses of science for defensive preparations in peace time and for other national interests, or with the training of recruits for research by grants of public money to the Universities. I think that it can properly be claimed that we knew here, even before the belief attracted a wider support and conviction, that a modern nation, as certainly as a well-organized modern industry, depends for success upon generous and far-seeing expenditure on scientific research and on the recruitment of first-rate ability to its service, and risks failure and disaster by parsimony and a narrow vision of its responsibilities in these directions

The response of our national scientific reserves to the demands of war might suggest that our national deficiency hitherto has been chiefly in provision for the applications of science, and that fundamental researches in this country have lacked less in opportunity and encouragement. Here too, however, if we make comparison with other countries, I think that we shall be obliged to conclude that our discoverers, as great as any, in a world era of great discovery, have often had to do their work in spite of a paucity of equipment and accommodation which would hardly have been conceivable elsewhere. That is a difficulty and a disparity which, unless some action is taken, will certainly increase with the growing demands of fundamental researches for elaborate and costly items of equipment. Recent discoveries have made such items available and necessary, at a rate which the war-time concentration of research will even have accelerated. Money to procure and to instal them will become ever more essential to work on the general front of scientific progress. The Society's Council, having considered a Memorandum by certain of its Fellows on the prospective needs of fundamental researches in physics, have already appointed a Committee to consider the position in detail. It seems unlikely that other departments of fundamental science, when news of this decision reaches their representatives, will wish the Society to assume that the needs of physics, even if more obvious than some others, are unique in their importance and urgency. With so many interests and authorities now directly concerned for applied science, we can hardly doubt, indeed, that it is to-day a primary duty and mission of the Royal Society, as of the related Societies having more special and restricted aims in science, to aid and to encourage researches which seek the advancement of knowledge without immediate reference to its use, though with a clear conviction that such progress is often a necessary condition of practical advance, or even the most direct way towards it. Care for the practical fruits of the tree of knowledge

was never, indeed, so urgent as to-day; but the tree will wither unless we take care also that the roots have nourishment and room for spreading.

The mention of this last necessity brings me to the problem of the need of the Royal Society, and the more pressing need of one, at least, of our neighbour societies here, for accommodation more worthy of their national importance. The matter has more than once been under discussion in the past, and has recently been a matter of renewed concern to the Society's officers. With the efficient help of our Assistant Secretary, Mr Griffith Davies, we have been making, in this connexion, a survey of the records dealing with the different homes of the Royal Society, from its foundation down to recent years. I think that on another occasion, when more time is available, the Society may like to hear a review in greater detail of this aspect of our history, which has many points of interest. To-day the mention of a few of these must suffice



The meeting room of the Royal Society in Somerset House

Early in our career the interest of the Crown in the Society, and perhaps a recognition of a duty to provide accommodation for it, were signified by King Charles the Second's grant of Chelsea College and its estates, as set forth in our third Charter of 1669 The property proved, alas, for various reasons, to be much more a burden than an asset, and Christopher Wren, in 1682, with the Council's approval and recorded gratitude, sold it back to the King for £1300. Meanwhile an opportunity for the Society to acquire a house of its own, built to the designs of Hooke and Wren, on a piece of land granted by Henry Howard from the grounds

of Arundel House, had not become effective. The Society, therefore, save for interruptions due to the plague and the fire of London, remained for fifty years from its foundation a tenant of rooms in Gresham College, till in 1710, when Isaac Newton was President, it acquired the house in Crane Court, off Fleet Street, which was its home for another sixty-eight years. In 1778, thanks to the personal interest in our affairs of King George III, the friend of our great President of those days, Sir Joseph Banks, the Society was granted quarters in Somerset House. Therewith the obligation of the state to provide us with housing was for the first time definitely accepted, 'in generous recognition by the Sovereign of the services which science had rendered to the state', as Banks stated in his address of 1780. The records show that the accommodation in Somerset House was regarded from the first as inadequate, even though our requirements had been reduced by the transfer of our 'Repository of Rarities' to the British Museum The rooms, on the other hand, can be seen from prints of the period to have had a pleasant dignity (see page 236), and the Society remained in them for nearly eighty years.

Towards the middle of the nineteenth century, a movement arose to secure new and better accommodation for the Royal Society, and at the same time for the other principal scientific societies then existing—the Linnaean, Geological, Astronomical and Chemical Societies. As early as 1847 a memorandum was under consideration by the newly founded Philosophical Club, a more seriously minded secession from, or rival to, the Royal Society Club of those days, formed under the same influences as those which had just carried the revised method of electing our Fellows. The Club presented proposals for bringing the major scientific societies under one roof, centralizing and co-ordinating their libraries without any attempt at fusion, providing three or four meeting rooms of different sizes, for use by the Societies in common and in turn, and, in general, making better provision for the interests common to all without any impairment of their independence in rules, traditions, procedures or property. When those of us who have been considering present-day needs look at this memorandum, as finally presented in 1852, we cannot but admire the foresight and wisdom of our mid-Victorian predecessors.

An opportunity of housing the scientific societies thus, as a community of co-operative but substantially independent units, was actually presented in the same year, 1852, by the offer of accommodation in new buildings then being planned on the estate at Kensington Gore, acquired with the proceeds of the 1851 Exhibition. We begin to see the benevolent interest of the Prince Consort in our concerns. Kensington Gore, we must remember, in those days of horse transport, was still on the rural margin of the suburbs, and, in gratefully declining the offer, the Royal Society and its associates urged upon the Government the desirability of housing the scientific societies centrally and, if possible, under a single roof. The acquisition by the Government, some years earlier, of Burlington House and its grounds, extending from the Piccadilly frontage through to the street which is now named Burlington Gardens, seemed, indeed, to have provided the ideal opportunity for giving effect to such a plan. The Prince Consort, with a vision of the future meaning

of science far in advance of his time, privately urged the five scientific societies to press their claim to the site. It had been understood, indeed, that the primary intention of the Government in buying Burlington House had been to provide accommodation for the learned societies. Lord Wrottesley, then our President, personally canvassed the Government, making it clear 'that the desire of the chartered societies for juxtaposition and for the Burlington House site was unabated'. Failing that, he indicated, they would be glad to be lodged in the buildings then occupied by the Royal Academy, that is, in what is now the National Gallery. The existence of a rival claim had become clear, and had, indeed, been mentioned to the Royal Society by the Prince Consort It appears that the Government had already made some kind of commitment to the Royal Academy, so far as the



Burlington House in 1868

mansion of Burlington House was concerned. It would take much too long to discuss even what is known of the rival lobbyings of those days. It must suffice for this occasion to recall the results, and to lament the fact that a magnificent opportunity was lost, which would have given London a scientific centre worthy of the nation's achievement. We cannot blame our predecessors, who probably did all that was possible, nor can we grudge their success to our friends of the Royal Academy, who are in no way to blame for taking what was offered to them. If the Government, indeed, had then used Burlington House and its grounds to discharge only these two of the obligations to which they were to some degree committed, the needs both of the scientific societies and of the Royal Academy could still have been handsomely met, and adequate scope for future development could have been ensured to both. The mansion itself with the wings of this front courtyard, already scheduled for rebuilding, could, for example, have been allotted to one, while the other of the two claimants could have had a new building, with frontage at the

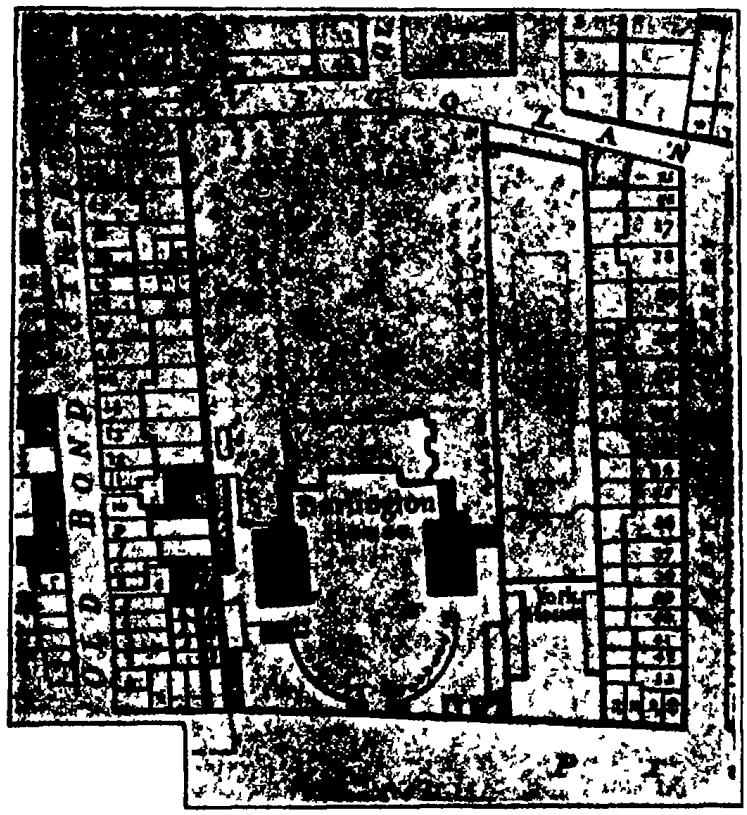
north end of the gardens, and ample space for extension southwards over them, to meet increasing needs and new developments (see plan, page 240). The Government, however, used Burlington House first to satisfy a third obligation which it had accepted, to house the University of London, then only a degree-giving body requiring space chiefly for the periodical examination of large numbers of candidates. Then in 1858, the continued pressure of the scientific societies and the Government's own desire to recover the rooms in Somerset House, led them to offer the use of Burlington House to the Royal Society, subject to the condition that for the time, and pending rebuilding on the sites round this courtyard, the Linnaean and Chemical Societies should be accommodated with us in the mansion, and that the University of London should still be able to use the large rooms in it for examinations. It is curious to reflect that this temporary arrangement gave to the Royal Society and its associates their only opportunity, to this day, even to share the use of a room suitable for a meeting of more than very modest dimensions.

In old Burlington House, then, we were established, and were to remain there with the Chemical and Linnaean Societies for some fifteen years; and an appearance of stability had at first been given to our occupancy by the mention of plans to build a new examination hall for the University of London on the western side of this quadrangle, and to allow the Royal Society to use this also for large meetings and for its gallery of portraits. Two later developments, however, dispelled any such hopes. In 1867 evidence came to the Society, first through a statement in *The Times*, that the Government had decided, after all, to give the Royal Academy a permanent lease of Burlington House and the right to extend northwards by building over its gardens. At about the same time, and presumably in fulfilment of another commitment, the large building which now fronts on to Burlington Gardens was begun, to accommodate the University of London and its examinations, and was opened by Queen Victoria in 1870.

The scientific societies were not, indeed, to be homeless, but the only possibility now left was to accommodate them in the buildings planned to be erected round this front courtyard, where they have been ever since. The total space thus offered did, indeed, allow more room to each of the societies than it had previously enjoyed, even after the Society of Antiquaries, at the Royal Society's instance, had been included in the scheme. But the space now available could not easily be planned for the sharing of meeting rooms and general facilities, or for a central federation of the libraries, or for any of the features of the earlier plan which would have enabled the societies to function as independent members of a real scientific community. The scheme had an even more fatal defect. The plans were made on the assumption that the societies existing in the 1860's, with their respective dimensions and requirements at that date, would provide a pattern of the needs of science for all time, or at least for the life of buildings designed to mid-Victorian standards of permanence. Each of these societies, therefore, with the approval of our own we must admit, presented its separate claim and had it embodied in the solidity of the buildings we still inhabit, filling the available space completely and precluding any

later expansion, rearrangement, or new admission to the circle thus finally closed. Societies which have changed but little in numbers or activities may have had little reason, even yet, to complain of the accommodation which they then acquired. For others the allotment, which had been regarded then as satisfying future needs for half a century at least, became obviously inadequate very much earlier.

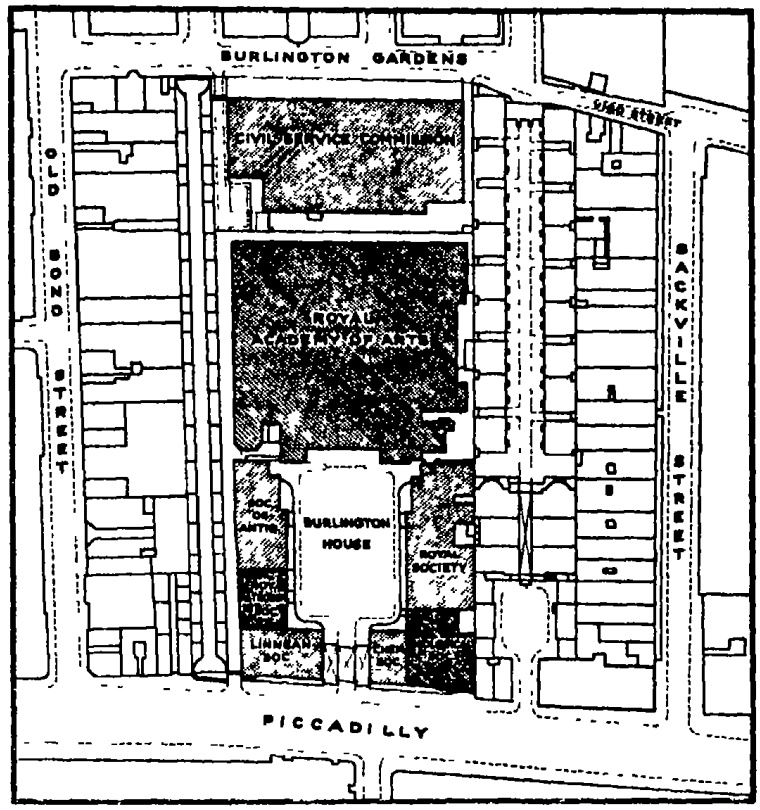
In 1900 came news that the large building on the Burlington Gardens frontage was to be vacated by London University, and tentative inquiry was immediately made as to the possibility of allotting it to the Royal Society, on the ground that



Plan of Burlington House in 1790

'the present rooms occupied by the Society were rapidly becoming inadequate'. The Government, however, had already decided to transfer the building to their Civil Service Commissioners, and it has continued to be dedicated to its original use for large-scale examinations, save for the later assignment of certain rooms in it to our much younger sister, the British Academy. It will be noted that the Royal Society was finding its quarters here inadequate as early as 1900, twenty-seven years after it entered them, and before there was even any prospect of the great expansion of its responsibilities and activities in recent years. Our accommodation is still the same to-day, after seventy years. Our walls cannot find room to hang our important collection of scientific portraits, and our great library is badly overcrowded, even though we have parted with some of it to give better housing, for a

time, to the remainder; and it continues, of course, to grow. Library pressure, in fact, is felt to varying degrees by all the societies here; and I think that it is still true, as some of our predecessors saw already in the 1850's, that no scheme will be able to deal with this problem efficiently, and to meet modern needs without disturbing historic associations, which does not include some kind of central co-ordination of libraries. The lack of a lecture or conference room, available in common for larger meetings, and well equipped with modern resources for projection and demonstration, is another acutely felt need. There are greater needs and



Plan of Burlington House in 1943

anomalies, however, than any of these common ones Of all the societies here the Chemical Society, which was originally satisfied with the poorest allotment of rooms, has undergone the greatest expansion In the 1860's it had a membership of some 450, it now has about 5000. Its library, of great importance to all workers in chemistry, whether fundamental or applied, has so burst the bounds of its accommodation, that a part of it is deposited in the crypt of a neighbouring church, and the Chemical Society's meeting room is in every way unsuitable, and inadequate to the meanest conception of the regular needs of a society of its standing and numbers. Apparently our predecessors of the 1870's did not see much future for chemistry. On the same evidence, they did not foresee any future for physics at all. The Physical Society did not then exist; by the time it was born there was no room

for its admission, and the State has never offered it a home. The same is true of other societies formed later to deal with functional aspects of biology and other new fields of knowledge. For most of their meetings these newer societies use and need the facilities available in academic and research institutions. A national centre of science, however, should be capable of progressive adjustment to changing needs, and we ought to be able to make new admissions, on a varying scale of allotment, to the central community of societies.

What then should we be doing to deal with the situation? Actions and decisions long past have imposed it upon us, and regrets and repinings over an opportunity lost more than seventy years ago will not help us to-day. We must admit, too, that our present quarters, with all their defects of elasticity, have provided a combination of central position with freedom from noise of traffic which might be hard to find again. Let me say, then, that the Royal Society's officers, having consulted with the officers of other societies here, and particularly of those whose needs are urgent or whose interests might be directly concerned, have not yet abandoned the attempt to find a solution which would not involve the removal of any from the Burlington House estate. If we fail in that direction—and there is no ground for optimism—the problem will remain, and the time is not one for neglect or postponement of action. On all hands we hear talk of reconstruction and see plans for the rebuilding of London. We cannot expect another Christopher Wren—one of our original Fellows and a leader in the science of his day, before he became its greatest architect; London missed that opportunity. It is natural and proper for the plans now being presented to make spacious and impressive provision in the new London for opera, drama, music and all the fine arts, and we shall surely join in a general welcome to any practicable scheme which can open the doors more widely to such cultural privileges, and enhance their dignity and worth in the eyes of London and of the nation. But I do not think that we must stand by and allow the claims of science again to go by default. A fear of overstatement, a passion for critical accuracy which is a part of the very spirit of science, may make us reluctant advocates. If necessary, however, we must be ready to remind all who may be concerned of the part which the British scientific effort has played, in making it possible now to plan at all, with confidence, for our own civic and national reconstruction. But for science, we may remind them, the very different plans which our enemies were so recently making for our future might already be taking effect. I have no doubt that the claim will be handsomely admitted, but we ought not to be too easily appeased with compliments and oratorical bouquets. The nation's opportunity, when peace returns, of enjoying the arts and the amenities of life will be dependent on its standards of health and prosperity, and these, in turn, ever more directly on science and its applications, as certainly as these are still needed to secure our national survival and victory in this war.

This ancient Royal Society of London, and those societies which have grown from it and round it in later years, constitute a scientific organism which is a national and imperial heritage, second to none in the world's esteem. Here are the

roots of the spreading tree of science and technology, which should form a major component of our national contribution to the new world now in the making. Seventy years ago these roots were given only enough soil for the replanting then undertaken; they have long been badly pot-bound, and some parts of the root system are threatened with strangulation, while others have appeared outside the pot. We can properly claim, I think, that the progressive needs of our scientific societies shall be given early consideration in any new allotment which plans for reconstruction may allow. We ought to have a scientific centre permitting them to co-ordinate their activities with economy, and giving room for change, expansion and organic growth by budding and division, in accordance with nature's law. I think that we have the further right to expect that the home of science, in this capital city, will have a dignity symbolizing its value to the nation and the empire, and enabling us to hold up our heads in the company of other countries, whose scientific academies, not more famous than ours, have so long been housed more worthily, and with a more generous recognition of their due place in an enlightened people's scale of cultural values

Matrix theory of correlations in a lattice. Part I

By R. Eisenschitz

(*Communicated by Sir Robert Robertson, F.R.S.—Received 5 March 1942—Revised 25 May 1943*)

The statistical mechanics of some crystalline systems may be reduced to statistical correlations between objects which are the unit cells of a fictitious lattice. The correlations are deduced from postulates according to which some configurations of the cells are incompatible with some configurations of the neighbouring cells, if, on the other hand, configurations of neighbours are compatible with each other, their probabilities are to combine by multiplication. By these postulates matrices are implicitly defined such that the probability distribution for a chain of cells is found by forming the powers of a matrix. A similar approach to the statistics of a lattice involves infinite matrices. It does not seem practicable to give explicit expressions for these matrices. If appropriate conditions are complied with, the correlations in a chain are accounted for by adjusting the mean probability coefficients of the cells and for the rest regarding the cells as statistically independent. In this case the infinite matrices may be replaced by the outer power of finite matrices. As result an equation is given by means of which the thermodynamical energy may be calculated as function of temperature.

General introduction

The statistical mechanics of systems involving correlations becomes increasingly important in the theory of solids and liquids. The rigorous treatment of the theory presents considerable difficulties. It seems advisable to treat the matter in such a way as to separate the mathematical formalism from an application of the theory to a concrete physical system. This is the course adopted here.

Part I of the paper deals with the mathematical aspect and in Part II the theory is given for a physical system.

1 Introduction to Part I

The present paper deals with the statistical mechanics of crystals. The forces in the crystal are assumed to be of such a kind that one atom interacts with only a finite number of other atoms in the lattice. A further assumption is made on the number of configurations of the crystal; while this number is infinite in an infinite crystal, it is assumed that its ratio to the number of lattice points is finite.

These assumptions are in fact never strictly realized; they are, however, frequently used as approximations. In mixed crystals, for instance, the partition function may adequately be separated into a vibrational and configurational factor, and the energy of configuration may be assumed as equal to the sum of bond energies (these are not necessarily restricted to bonds of pairs or interaction of nearest neighbours). Some fair approximation may be expected in crystals with rotating molecules by substituting a finite number for the infinity of molecular orientations. In ferromagnetism a rough approximation may be obtained by considering the spin interaction of

nearest neighbours and neglecting the resonance of states with equal magnetic moment.

The theory of the above systems involves statistical correlations between different atoms in the lattice which may extend over long distances.

In this paper a theory is given which applies to a large section of these correlation problems It is presented without reference to any special physical system Lassettre & Hove (1941) and Montroll (1941) have made important contributions to the theory The main result of this paper is to show how the statistics of the crystal can be reduced to a statistics of independent systems This has apparently not yet been anticipated generally but has been recognized by the author (1941) in a special case.

2. Mathematical groundwork

A few results of the theory of probability required in this investigation are obtained separately in this section of the paper.

Consider now objects which may occupy configurations $c_1 \dots c_\lambda \dots c_A$ with the (not necessarily normalized) probability coefficients $p_1 \dots p_\lambda \dots p_A$ In a collection of these objects configurations are specified by the configurations c_λ of the particular objects. In a configuration of the collection the 'populations of the configurations c_λ' are defined as the number of objects which occupy the configurations c_λ. The objects are said to be statistically independent if they occupy their configurations with a probability that is independent of the configurations which the other objects of the collection may occupy Otherwise there are 'statistical correlations' in the collection A collection of statistically independent objects will be called an 'assembly'; a collection of statistically dependent objects is to be spoken of as an 'aggregate'.

There are well-known theorems on probability which hold for any assembly but not generally for aggregates. In an assembly of independent aggregates relations may be expected to hold which are in some way independent of the special kind of correlation within the aggregates Some results of this kind will be obtained in the following.

(*a*) First consider the statistics of independent objects In an assembly of n objects let the populations of the configurations be denoted by $nr_1 \quad nr_\lambda \quad nr_A$, r_λ are positive fractions subject to the condition $\sum_\lambda r_\lambda = 1$ All configurations of the assembly which are not distinct by their populations nr_λ have equal probability. The total probability of the configurations with assigned populations is equal to a term in the expansion of

$$F(n, p) = (p_1 + p_2 + \dots + p_A)^n \tag{1a}$$

Regarding $F(n, p)$ as a function of $p_1, p_2, \dots, p_A$, it will be spoken of as 'function of distribution' Denoting the terms in the expansion by

$$w(nr_1, nr_2, \dots, nr_A, p_1, p_2, \dots, p_A) = \omega(nr_1, nr_2, \dots, nr_A)\, p_1^{nr_1} p_2^{nr_2} \dots p_A^{nr_A},$$

briefly by $$w(nr, p) = \omega(nr)\, \Pi_\lambda p_\lambda^{nr_\lambda},$$

it is seen that $$\omega(nr) = n!/\Pi_\lambda(nr_\lambda!).$$

In the limit $n \to \infty$ the polynomial (1a) has a greatest term

$$n!\, \Pi_\lambda p_\lambda^{n\bar{r}\lambda}/\Pi_\lambda (n\bar{r}_\lambda!),$$

where $\bar{r}_\lambda = p_\lambda/\sum_\lambda p_\lambda$, in comparison with which the remaining terms may be neglected. In this limit

$$w(nr, p) \doteq \exp\left(-\frac{n}{2}\sum_\lambda (r_\lambda - \bar{r}_\lambda)^2/\bar{r}_\lambda\right).$$

(b) In an aggregate of k objects the function of distribution does no more comply with equation (1a) but is to be written

$$F(k, p) = \sum_{s_1}\sum_{s_2} \ldots \sum_{s_\lambda}\sum_{\tau} \omega_\tau(ks_1, ks_2, \ldots, ks_\Lambda)\, p_1^{ks_1} p_2^{ks_2} \ldots p_\Lambda^{ks_\Lambda}, \tag{1b}$$

where ks_λ is the population of the configuration c_λ, the suffix τ refers to all possible permutations of objects occupying different configurations, and $\omega_\tau(ks_1, ks_2, \ldots, ks_\Lambda)$ are arbitrary non-negative numbers. Of course $\sum_\lambda s_\lambda = 1$. The average populations $k\bar{s}_\lambda$ are

$$k\bar{s}_\lambda = p_\lambda\, \partial \log F(k, p)/\partial p_\lambda. \tag{2}$$

The mean probability with which one object in the aggregate occupies the configuration c_λ is different from p_λ and equal to p'_λ and

$$\bar{s}_\lambda = p'_\lambda/\sum_\lambda p'_\lambda. \tag{3}$$

Now consider an assembly of m independent aggregates containing $mk = n$ objects. The populations in the assembly are denoted by nr_λ, the coefficients of total probability for the configurations with assigned populations are denoted by

$$w(nr_1 \ldots nr_\Lambda, p_1 \ldots p_\Lambda).$$

In the limit $m \to \infty$ these probabilities are given by the same expression as apply to an assembly of independent objects, namely,

$$w(nr, p) \doteq \exp\left(-\frac{n}{2}\sum_\lambda (r_\lambda - \bar{r}_\lambda)^2/\bar{r}_\lambda\right), \tag{4}$$

where obviously $$\bar{r}_\lambda = \bar{s}_\lambda. \tag{5}$$

In order to prove this, the function of distribution for the aggregate is written $F(k, p) = \sum_\kappa P_\kappa$, $\kappa = 1 \ldots K$, where the configurations of the aggregate are enumerated by the suffix κ. The populations are denoted by $ks_{\kappa\lambda}$. $\sum_\lambda s_{\kappa\lambda} = 1$.

The function of distribution of the assembly is

$$F(n, p) = [F(k, p)]^m \tag{6}$$

and

$$F(n, p) = \sum_{l_1} \ldots \sum_{l_K} m!\, \Pi_\kappa P_\kappa^{ml_\kappa}/\Pi_\kappa (ml_\kappa!), \tag{6a}$$

where $\sum_{\kappa} l_{\kappa} = 1$ In the limit $m \to \infty$

$$F(n, p) \doteq \sum_{l_1} \cdots \sum_{l_K} \exp\left(-\frac{m}{2} \sum_{\kappa} (l_\kappa - \bar{l}_\kappa)^2/\bar{l}_\kappa\right) \tag{6b}$$

The numbers r_λ and l_κ are connected by the equations

$$r_\lambda = \sum_{\kappa} l_\kappa s_{\kappa\lambda}. \tag{7}$$

The sums in equation (6*b*) may be divided into partial sums in each of which all terms have equal populations $nr_1, nr_2, \ldots, nr_\Lambda$, every partial sum is the function of distribution for the configurations having these populations. This function of distribution may be alternatively expressed by means of Λ parameters $\beta_1, \beta_2, \ldots, \beta_\Lambda$.

$$\sum_{l_1} \sum_{l_2} \cdots \sum_{l_K} \exp\left(-\frac{m}{2} \sum_{\kappa} \left\{[(l_\kappa - \bar{l}_\kappa)^2/\bar{l}_\kappa] + 2k \sum_{\lambda} l_\kappa s_{\kappa\lambda}\beta_\lambda\right\}\right).$$

Rearranging this expression into

$$\sum_{l_1} \sum_{l_2} \cdots \sum_{l_K} \exp\left(-\frac{m}{2} \sum_{\kappa} \left\{[l_\kappa - \bar{l}_\kappa(1 - k\sum_{\lambda} s_{\kappa\lambda}\beta_\lambda)]^2/\bar{l}_\kappa + 2\bar{l}_\kappa k(\sum_{\lambda} s_{\kappa\lambda}\beta_\lambda) - \bar{l}_\kappa k^2(\sum_{\lambda} s_{\kappa\lambda}\beta_\lambda)^2\right\}\right),$$

it is seen to approach asymptotically its greatest term in which $l_\kappa = \bar{l}_\kappa$ and where

$$\bar{l}_\kappa = \bar{l}_\kappa(1 - k\sum_{\lambda} s_{\kappa\lambda}\beta_\lambda). \tag{8}$$

It is accordingly allowed to substitute on the right of (6*b*) the greatest term for each partial sum so that

$$(\sum_{\kappa} P_\kappa)^m \doteq \sum_{l_1} \cdots \sum_{l_K} \exp\left(-\frac{m}{2} \sum_{\kappa} (l_\kappa - \bar{l})^2/\bar{l}_\kappa\right).$$

In the latter sum every term is equal to the total probability of the configurations of the assembly with assigned values of the numbers $\bar{l}_\kappa$.

By putting $l_\kappa = \bar{l}_\kappa$ in equations (7) and noting that

$$\sum_{\kappa,\lambda} \bar{l}_\kappa s_{\kappa\lambda}\beta_\lambda = 0$$

—which equation is derived from (8)—Λ independent linear equations are provided for calculating the parameters β_λ It follows that these parameters and accordingly the numbers $\bar{l}_\kappa$ are linear functions of the numbers r_λ and

$$(\sum_{\kappa} P_\kappa)^m \doteq \sum_{r_1} \sum_{r_2} \cdots \sum_{r_\Lambda} \exp\left(-\frac{m}{2} Q(r_\lambda)\right),$$

where $Q(r_\lambda)$ is a polynomial of second degree in the variables r_λ The summation is carried out over those systems of values of $r_1, r_2, \ldots, r_\Lambda$ to which a non-negligible probability corresponds.

Any function of distribution of this type contains no other constants than those which are functions of the moments of the first and second order, namely, $\bar{r}_\lambda$, $\overline{r_\lambda^2}$,

$(\overline{r_\lambda r_{\lambda'}})$. The linear moments are proportional to the mean probability coefficients of one object as given in equations (5) and (3). The moments of the second order depend similarly upon the mean probability coefficients of pairs of objects. Denoting by $p'_{\lambda\lambda}$ the mean probability coefficient for two objects to occupy simultaneously the configuration c_λ, then

$$p'_{\lambda\lambda} = (n\overline{r_\lambda^2} - \bar{r}_\lambda)/(n-1).$$

Denoting by $p'_{\lambda\lambda'}$ the mean probability coefficient for one object to occupy the configuration c_λ and another object to occupy simultaneously the configuration $c_{\lambda'}$, then

$$p'_{\lambda\lambda'} = 2n\overline{(r_\lambda r_{\lambda'})}/(n-1).$$

The probabilities $p'_{\lambda\lambda}$ and $p'_{\lambda\lambda'}$ are averages over all pairs of objects. There are $n(n-1)/2$ pairs The number of those pairs the objects of which are contained in different aggregates is $m(m-1)\,k^2/2$ which is asymptotically equal to the former number. The overwhelming majority are accordingly pairs of statistically independent objects. The moments of the second order and the function $Q(r_\lambda)$ are therefore equal to the corresponding quantities in an assembly of independent objects which occupy the configurations c_λ with the probability p'_λ; $Q(r_\lambda) = \sum_\lambda (r_\lambda - \bar{r}_\lambda)^2/\bar{r}_\lambda$ such as asserted in equation (4).

(c) The configurations c_λ may be divided into classes to be denoted by C_μ; populations of the classes C_μ may be defined by the sum of populations of all those configurations c_λ which belong to the class C_μ. The populations of the classes C_μ are denoted by kS_μ; S_μ is the sum of some s_λ and thus

$$S_\mu = \sum_{\lambda(\mu)} s_{\lambda(\mu)}, \tag{9}$$

where the suffix in brackets indicates the class C_μ to which the corresponding configuration c_λ belongs to

The total probability for the configurations of the aggregate with assigned values of the S_μ and the function of distribution for these configurations is denoted by $W(kS_\mu, p)$ and

$$W(kS_\mu, p) = \sum_{s_{1(\mu)}} \sum_{s_{2(\mu)}} \cdots \sum_{s_{\Lambda(\mu)}} w(ks_{1(\mu)}, ks_{2(\mu)}, \ldots, ks_{\Lambda(\mu)}).$$

The sum is to be taken over all those values of $s_{1(\mu)}, s_{2(\mu)}, \ldots, s_{\Lambda(\mu)}$ which satisfy equation (9) for the assigned values of the S_μ.

In an assembly of m independent aggregates the populations of the classes C_μ are denoted by nR_μ, $n = mk$. The function of distribution for the configurations with assigned populations of the classes C_μ is denoted by $W(nR_\mu, p)$.

Obviously

$$F(k, p) = \sum_{S_\mu} W(kS_\mu, p)$$

and

$$F(n, p) = \sum_{R_\mu} W(nR_\mu, p),$$

and according to equation (6)

$$\sum_{R_\mu} W(nR_\mu, p) = \left[\sum_{S_\mu} W(kS_\mu, p)\right]^m. \tag{10}$$

One term on the left of (10) is equal to a partial sum of terms on the right. By substituting for the partial sum its greatest term, $W(nR_\mu, p)$ is obtained as a product of powers of functions $W(kS_\mu, p)$ multiplied by a factor which is independent of $p_1, p_2, \ldots, p_A$ but may depend upon $R_1, R_2, \ldots, R_\mu, \ldots$. If there is in $W(kS_\mu, p)$ one term in comparison with which the remaining terms may be neglected, the product of powers is reduced to $[W(kS_\mu, p)]^m$ with $S_\mu = R_\mu$.

(*d*) By assigning to every object in an aggregate a class C_μ such that the object may occupy no other configuration than those belonging to C_μ, a set of configurations of the aggregate is selected. In this set all configurations have the same populations of the classes $C_1, C_2, \ldots, C_\mu, \ldots$ and therefore are specified by the same values of the numbers $S_1, S_2, \ldots, S_\mu, \ldots$. By permuting the objects to which different classes C_μ are assigned, another set of configurations is obtained in which the numbers S_μ necessarily have the same values as in the first set. There are $k!/\Pi_\mu (kS_\mu!)$ different sets in which the numbers $S_1, S_2, \ldots, S_\mu, \ldots$ have the same values.

The function of distribution of a set depends necessarily upon the numbers $S_1, S_2, \ldots, S_\mu, \ldots$ and possibly varies with the permutations. It is accordingly denoted by $v_\sigma(kS_\mu, p)$, the suffix σ ranging from 1 to $k!/\Pi_\mu (kS_\mu!)$. There may be sets for which this function is equal to 0. Obviously

$$\sum_\sigma v_\sigma(kS_\mu, p) = W(kS_\mu, p). \tag{11a}$$

In an assembly of independent aggregates similar sets of configurations are selected by assigning a class C'_μ to every object. Their function of distribution is denoted by $v_\rho(nR_\mu, p)$, the suffix ρ ranging from 1 to $n!/\Pi_\mu (nR_\mu!)$. Obviously

$$\sum_\rho v_\rho(nR_\mu, p) = W'(nR_\mu, p) \tag{11b}$$

Given the values of $R_1, R_2, \ldots, R_\mu, \ldots$, then those corresponding functions $v_\rho(nR_\mu, p)$, which are different from 0 and not negligibly small, approach asymptotically ($\lim m \to \infty$) equality among themselves and with their average which is proportional to $W(nR_\mu, p)$.

In proving this I assume for sake of argument that given the values of $S_1, S_2, \ldots, S_\mu, \ldots$, the functions $v_\sigma(kS_\mu, p)$ are distinct from each other, otherwise the assertion would be trivial.

It follows from equations (10), (11*a*), (11*b*) that every function $v_\rho(nR_\mu, p)$ is a sum of products of powers of functions $v_\sigma(kS_\mu, p)$. In order to prove the assertion it is sufficient to show that the sum of power products is asymptotically reduced to one term. Assuming at first that

$$W(nR_\mu, p) = \gamma[W(kS_\mu, p)]^m, \tag{12}$$

where γ does not depend upon the p_λ but possibly upon the R_μ and where $R_\mu = S_\mu$, then

$$W(nR_\mu, p) = \gamma[\Sigma v_\sigma(kS_\mu, p)]^m. \tag{13}$$

Every term in the expansion of the right is equal to a sum of equal functions $v_\rho(nR_\mu, p)$, and on the other hand a product of high powers of functions $v_\sigma(kS_\mu, p)$. The expansion is asymptotically equal to one of its terms. A similar argument applies if the right of equation (12) is a product of powers of the $W(kS_\mu, p)$.

The factor γ depends upon the number of functions which deviate appreciably from the average. γ is to be regarded as an unknown continuous function of the numbers R_μ.

With regard to those functions $v_\mu(nR_\mu, p)$ which are not negligibly small, it is found finally that

$$v_\rho(nR_\mu, p) = \alpha' \; W(nR_\mu, p), \tag{14}$$

where α' is an unknown continuous function of the R_μ. Equations (4) and (14) are part of the necessary foundations of the subsequent calculations.

3. Formulation of the statistical problem

Since every atom interacts with only a finite number of other atoms, it is possible to divide the lattice into equal domains of such a kind that every atom interacts only with those atoms which are contained in the same domain. These domains overlap, so that a lattice point is generally contained in more than one domain. From a geometrical point of view the lattice may be said to be generated by a group of three displacements of the unit cells or of the domains

In a lattice in which there is interaction between nearest neighbours only, the number of atoms divided by the number of domains in the crystal is equal to the number of atoms in the unit cell; the atoms which are sited on the boundary of the unit cells belong to at least two domains between which they are shared. If atomic interaction extends further, the above ratio is larger than the number of atoms in the unit cell and the domains overlap to greater depth.

The configurations of the domain are specified by the configurations of the lattice points. The latter are specified by the kind of atom which occupies the lattice point, the internal state of the atom, its displacement from equilibrium position, etc. Since the configuration of every lattice point enters into the specification of the configuration of the domain, I shall speak of the 'contribution' of the lattice point to the configuration of the domain, although no quantity expressible in numbers is contributed. The energy of the domain is, on the other hand, equal to the sum of contributions of energy made by atoms and bonds between atoms. In defining the energy of the domain each of these contributions is divided by the number of domains between which the atoms or bonds, respectively, are shared. The energy of the lattice is consequently obtained as the sum of energies of domains.

Mutually overlapping domains may not occupy their configurations independently of each other, for each lattice point which is shared by the domains makes the same contribution whether it is regarded to belong to one domain or the other.

In any pair of domains there are lattice points corresponding to each other, such that the set of distance vectors from one point in one domain to the remaining points of the same domain is equal to the set of distance vectors from the corresponding point in the other domain to the remaining points of the other domain. The distance between corresponding points in different domains defines the distance between the domains. It is possible to select from those domains, with which one particular domain overlaps, six distinct domains in three non-coplanar directions with the three smallest distances. These six domains are said to be 'neighbours' of the original domain. A frame of rectilinear co-ordinates x, y, z is defined such that the axes are parallel to the three vectors joining neighbours. To every domain three integer co-ordinates are assigned in relation to these axes

In an isolated domain a probability coefficient is assigned to every configuration. In neighbouring domains some configurations of one neighbour are incompatible with some configurations of the other If two configurations combine, the probability coefficient of the combination is equal to the product of the two probability coefficients corresponding to either domain, this follows from the additivity of domain energy.

It is possible to link any pair of two domains in the crystal by a series of neighbouring domains. There are accordingly statistical correlations between all domains of the lattice.

The probability coefficients of the configurations of the lattice may accordingly be derived from the probability coefficients of the domains provided the correlations between neighbours are correctly taken into account. In the terminology of § 2 the lattice may be regarded as an aggregate of domains, the statistics of which is determined by correlations between neighbouring domains. The objects of this aggregate are specified by enumerating their configurations individually (i.e. the configurations of the domains) and by establishing which of the objects are neighbours. The correlations between neighbours have to be formulated as 'rules of composition'

The configurations of the domains may obviously be distinguished from each other and their number is finite, therefore a symbol may be assigned to each of them, namely, c_1, c_2, .., c_λ, .., c_A Let E_λ be the energy of the configuration c_λ, $E_\lambda \geqslant 0$. The corresponding probability coefficient is $p_\lambda = \exp(-E_\lambda/kT) = t^{E_\lambda}$. Excluding absolute zero of temperature it is assumed that $0 < t \leqslant 1$.

Every object of which the aggregate is composed represents one particular domain. The object is accordingly labelled by means of the co-ordinates of this domain. The objects are abstractions to which no position in space is assigned. It is nevertheless possible and convenient to speak of 'neighbouring' objects, which represent neighbouring domains, and to consider accordingly the totality of objects as forming a fictitious lattice. This expression is used in the following. The objects of the aggregate are considered to be the unit cells of the fictitious lattice. In order to distinguish them from the unit cells of the crystal lattice, these unit cells are spoken of as 'cells'. In relation to the co-ordinates of the domains we shall use expressions such as 'x-neighbours', 'z-chains', 'xy-layers', etc.

In order to formulate the rules of composition we consider first a pair of neighbouring domains which are represented by cells with the co-ordinates x, y, z and $x+1, y, z$. Whether or not configurations of these cells may combine depends upon the contribution of those lattice points which are shared by the domains. According to the contribution of these lattice points, the configurations are partitioned into classes such that either all configurations or no configurations of the class may combine with any arbitrary configuration of the neighbouring domain. All those configurations of the first domain to which the above lattice points make the same contribution are assembled in a class to be denoted by $B(x)_h$; $h = 1, \ldots, \Gamma$. All those configurations of the second domain to which the above lattice points make the same contribution are assembled in a class $A(x)_g$; $g = 1, \ldots, \Gamma$. The configurations of the first domain which are contained in one particular class combine with those configurations of the second domain contained in one particular class. The suffices g and h are chosen such that their values are equal for the classes of combining configurations.

Similarly, classes of combining configurations are defined for y-neighbours and z-neighbours.

The configurations c_λ are accordingly partitioned in six different ways into classes which are denoted by $A(x)_g$, $B(x)_h$, $A(y)_i$, $B(y)_j$, $A(z)_k$, $B(z)_l$. $g, h = 1, \ldots, \Gamma$, $i, j = 1, \ldots, \Delta$; $k, l = 1, \ldots, \Theta$. Every c_λ is contained in one and only one of the classes $A(x)_g$ and $B(x)_h$ and $A(y)_i$, etc. Considering a cell with the co-ordinates x, y, z, those configurations which are contained in the class $B(x)_g\{B(y)_i; B(z)_k\}$ combine with all those and no other configurations of the cell at $x+1, y, z$ $\{x, y+1, z; x, y, z+1\}$ which are contained in the class $A(x)_g\{A(y)_i; A(z)_k\}$.

If c_λ and $c_{\lambda'}$ are combining configurations of a pair of neighbouring cells, the corresponding configuration of the pair (denoted by $c_\lambda c_{\lambda'}$) has the relative probability $p_\lambda p_{\lambda'}$.

These are the premisses from which the probability coefficients of the aggregate of domains are to be deduced.

In dealing with this problem I employ 'functions of distribution' which are sums of relative probabilities taken over all configurations of a cell, chain, layer or the entire fictitious lattice. These functions are distinct from the corresponding partition functions for two reasons. (*a*) They are functions of the variables $p_1 \ldots p_\Lambda$, whereas the partition function is the function of one variable, namely, the temperature. Similarly, as the partition function provides the mean energy, these functions provide a more detailed information, namely, the mean probabilities with which the domains within a chain, etc., occupy their different configurations. Information of this kind is necessary for calculating the correlations between different chains and layers. The thermodynamic energy is obtained as average energy of the configurations of the domains. (*b*) The numerical values of the functions of distribution are generally distinct from the numerical values of the partition functions and their ratio may depend upon temperature. The thermodynamic energy cannot be obtained from the function of distribution by applying the usual formula by which that energy is related to the partition function.

4. Application of matrices

The above rule of composition for x-neighbours is readily expressed in terms of matrices. A matrix $D(x)$ of Γ rows and columns can be defined such that the matrix element $D(x)_{gh}$ contains the (symbolic) sum of all those c_λ which are simultaneously contained in the classes $A(x)_g$ and $B(x)_h$; if there is no c_λ which corresponds to a pair of suffices g and h, the matrix element $D(x)_{gh}$ is 0. According to the matrix law of multiplication, the matrix elements of $D(x)^2$ are sums of products $c_\lambda c_{\lambda'}$; all products of combining configurations and no other products appear in these matrix elements. By substituting p_λ for c_λ in $D(x)$, a matrix $G(x)$ is obtained, the elements of which are sums of probability coefficients. The sum of the matrix elements of $G(x)^2$ is equal to the sum of probability coefficients taken over all configurations of a pair of x-neighbours.

Similarly, matrices $D(y)$, $G(y)$ and $D(z)$, $G(z)$ can be defined, in terms of which the rules of composition for y-neighbours and z-neighbours are formulated.

The configurations of an x-chain and their probability coefficients are obviously found by calculating the nth power of the matrices $D(x)$ and $G(x)$ respectively. $\sum_{g,h} G(x)^n{}_{gh}$ is equal to the partition function (function of distribution respectively) of an x-chain.

An xy-layer is composed of neighbouring x-chains the cells of which combine according to the rule for y-neighbours. The layer may be regarded as a chain the cells of which are x-chains. The rule of composition for y-neighbours implies a rule of composition for neighbouring x-chains by which a matrix L is implicitly defined. As the number of configurations increases with the length of an x-chain, it may be impracticable to enumerate the individual configurations and to find expressions for the matrix elements of L. Actually it is not these matrix elements but the sum $\sum_{i,j} L^m{}_{ij}$ which is required. The evaluation of the sum—which is the function of distribution of the xy-layer—is the central problem of this paper and will be given in §§ 5 and 6. The lattice may similarly be regarded as composed of neighbouring xy-layers; a matrix L^+ is implicitly defined by the rule of composition for two neighbouring xy-layers.

In order to employ the powerful methods of matrix algebra it will be assumed that the matrices comply with certain conditions This involves a restriction of the scope of the theory.

At first it is assumed that the matrices $G(x)$, $G(y)$, $G(z)$ are symmetric matrices. If this condition holds, it may be shown that the matrices L and L^+ are also symmetric. There exists accordingly a real orthogonal transformation by which these matrices are transformed to diagonal form.

The significance of this symmetry is conveniently demonstrated by considering those lattice points of a domain which are shared by the right neighbour and those which are shared by the left neighbour (for instance, those represented by the two x-neighbours of a cell). Every lattice point of the first kind is geometrically equi-

valent to one lattice point of the second kind. By permuting the contributions of geometrically equivalent lattice points any configuration of the domain is transformed into another of its configurations. These two configurations have equal probability coefficients if and only if the corresponding matrix—in this instance the matrix $G(x)$—is symmetric. The symmetry of these matrices is accordingly not derived from crystal symmetry.

It will be assumed that some matrices have the following properties: Their highest proper value and the highest proper value of their matrix square are single and the latter is the square of the former. Matrices of this kind will be said to be 'G-matrices' and the proper vector corresponding to the highest proper value will be spoken of as 'first proper vector'.

Every real symmetric matrix M may be represented in terms of its proper values*

$$M_{ik} = \sum_j M'_j \xi_{ji} \xi_{jk}, \tag{15}$$

where M'_j are the proper values and ξ_{ji}, ξ_{jk} the components of the proper vectors. If M is a high power of a G-matrix, the sum in equation (15) is reduced to one term. In this case M is equal to a number (i e. the power of the highest proper value) multiplied by an idempotent matrix.

A matrix in which the sum of matrix elements is equal for all rows is a G-matrix and the components of the first proper vector are equal to each other. (The proof is simple and may be omitted) If it is assumed that the number of configurations c_λ is equal in every row of $D(x)$, it follows that at high temperatures ($t = 1$) the matrix $G(x)$ is a G-matrix and ξ_i the components of the first proper vector are equal to each other. There is accordingly a range $t \leqslant 1$ where $G(x)$ and the matrix $G(x)_{\sigma\lambda} \xi_\sigma \xi_\lambda$ are G-matrices

The assumption that there are equal numbers of c_λ in every row of $D(x)$ is not necessary for the validity of the following calculations. The assumption may hold for special systems of physical importance. It will, however, be assumed that the matrices $G(x)$, $G(y)$, $G(z)$, and some other matrices to be defined in the course of calculations, are G-matrices. By this condition the scope of the theory is once more restricted. In applying the theory to any special system it is necessary to prove that the above conditions are complied with

5. The statistics of a chain

As far as the energy is concerned, the probability distribution of an x-chain is determined by its partition function† which is equal to $\sum_{\sigma,\lambda} G(x)^{n_{\sigma\lambda}}$. The configurations of the x-chain have, however, to be specified not only by their energy but also with regard to those configurations of a neighbouring x-chain with which they combine.

* Cf. Temple (1934, p. 15).

† In the case of a chain a function of distribution will be defined which is numerically equal to the partition function at all temperatures.

It is, on the other hand, impracticable to enumerate the individual configurations of an x-chain.

If the x-chain was replaced by a collection of independent cells, the problem would be considerably simplified. It would be sufficient to specify every configuration of the collection by the appropriate populations of cell configurations irrespective of the position of the cells In a collection of independent cells no distinction is made between two different configurations in which $nr_1, nr_2, \ldots, nr_A$ cells occupy the configurations $c_1, c_2, \ldots, c_A$ respectively. The two configurations combine with two different sets of configurations of their y-neighbours, but these sets have equal functions of distribution.

It will be seen that there is a close similarity between the statistics of an x-chain and of a collection of statistically independent cells so that it is sufficient to specify the configurations of an x-chain by their populations of cell configurations.

In the terminology of § 2, an x-chain is an aggregate of cells. In applying equation (1 *b*) it is noticed that all configurations with the same set of populations c_λ have the same probability. The function of distribution may accordingly be written without the suffix τ

$$F(k, p) = \sum_{s_1}\sum_{s_2} \ldots \sum_{s_1} \omega(ks_1, ks_2, \ldots, ks_A)\, p_1^{ks_1} p_2^{ks_2} \ldots p_A^{ks_A}$$

$$= \sum_{s_1}\sum_{s_2} \ldots \sum_{s_A} w(ks_1, ks_2, \ldots, ks_A; p_1, p_2, \ldots, p_A),$$

where the probability coefficients ω are unknown functions of the populations.

The function of distribution is, on the other hand, equal to $\sum_{g,h} G(x)^{k_{gh}}$ provided that the matrix elements are regarded as functions of $p_1, p_2, \ldots, p_A$. Since $G(x)$ is supposed to be a G-matrix

$$F(k, p) = (G')^k \sum_{g,h} \xi_g \xi_h, \tag{16}$$

where G' is the highest proper value of $G(x)$ and ξ_g, ξ_h are the components of the first proper vector. In this function of distribution it is admissible to regard $\left(\sum_{g,h} \xi_g \xi_h\right)$ as virtually constant.

The function of distribution for an assembly of m independent chains is

$$(G')^{km}\left(\sum_{g,h} \xi_g \xi_h\right)^m,$$

and for a chain of km cells $\qquad (G')^{km} \sum_{g,h} \xi_g \xi_h.$

The difference between these functions divided by their average approaches 0 if k is large enough. For the purpose of calculating probabilities we may therefore put

$$\lim_{k\to\infty} \Sigma G^{km_{gh}} = \lim_{k\to\infty} [\Sigma G(x)^{k_{gh}}]^m. \tag{17}$$

Comparing (17) with (6) the x-chain is seen to have the statistics of an assembly of independent aggregates.*

According to equation (4) the probability coefficient for the configurations with assigned populations is equal to

$$w(nr_1, nr_2, \ldots, nr_\Lambda; p_1, p_2, \ldots, p_\Lambda) = \exp\left(-\frac{n}{2}\sum_\lambda (r_\lambda - \bar{r}_\lambda)^2/\bar{r}_\lambda\right),$$

where n is the number of cells and nr_λ are the populations of cell configurations. The values of r_λ could be found according to equation (2) by differentiation, but it is more convenient to consider a chain with a 'central' cell to which n and n' cells are joined to the left and right respectively. The function of distribution

$$\sum_{f,i}[G(x)^n\, G(x)\, G(x)^{n'}]_{fi} = \sum_{f,i}\sum_{g,h}\xi_f (G')^n\, \xi_g\, G(x)_{gh}\, \xi_h (G')^{n'}\, \xi_i$$

$$= (G')^{n+n'}\left[\sum_{f,i}\xi_f\xi_i\right]\cdot\sum_{g,h}G(x)_{gh}\,\xi_g\xi_h$$

is expanded in terms of the matrix elements of $G(x)_{gh}$ representing the central cell. The mean probability of the configuration c_λ is accordingly

$$p'_\lambda = p_\lambda \xi_g \xi_h, \tag{18}$$

where the suffices g, h correspond to the matrix element $D(x)_{gh}$ in which c_λ is contained. According to (3')

$$\bar{r}_\lambda = p'_\lambda / \sum_\lambda p'_\lambda$$

Consider next the functions of distribution of certain classes or sets of configurations which have a similar significance for the combination of chains as the classes $A(x)_i$, etc. have for the combination of cells.

In a chain of domains a set of configurations is selected when all lattice points vary their configurations with exception of those lattice points which are shared by a neighbouring chain. Any arbitrary configuration of the latter chain combines either with all configurations or with no configuration of the set.

Let the chain be represented by x-chains in the fictitious lattice which have the co-ordinates y and $y+1$ respectively, and c_λ be a configuration of the cell in the first

* It might be objected that G' is an irrational function of the p_λ and accordingly $(G')^{km}$ cannot be the distribution function of an assembly of aggregates which necessarily is a high power of a polynomial. $(G')^{km}$ approaches actually a high power of a polynomial. *Proof*: The diagonal sum of a matrix is equal to the sum of its proper values so that

$$(G')^k = \left[\sum_h G(x)^{k\lambda_h}\right](1+\delta_1),$$

where $|\delta_1|^{1/k} < 1$ and $|k\delta_1|$ is smaller than any given positive number if k is large enough. For any integer m, $0 < m < k$,

$$(G')^{mk} = \left[\sum_h G(x)^{k\lambda_h}\right]^m (1+\delta_2),$$

where $1+|\delta_2| \leqslant (1+|\delta_1|)^m$ and accordingly $|\delta_2| < 2k\delta_1$ may be held below any given positive number and m beyond any other given positive number if k is large enough.

chain. The contributions of those lattice points which are shared by the chains may be specified by the particular class $B(y)_j$ where the c_λ belongs to. The representative of the above set of configurations is accordingly selected by assigning to every cell of the first chain one of the classes $B(y)_j$ such that the cell may occupy no other c_λ than those belonging to $B(y)_j$. Let nR'_j be the number of cells to which the class $B(y)_j$ is assigned; $\sum_j R'_j = 1$. If the numbers $R'_1 \dots R'_r$ are given, there are many possible positions for the cells to which a certain class $B(y)_j$ is assigned so that there are $n!/\Pi_j(nR'_j!)$ different sets with assigned numbers R'_j.

The function of distribution of each set, a function of $p_1 \dots p_\lambda$, may depend on the numbers $R'_1 \dots R'_r$ and also on the positions of the cells with assigned classes $B(y)_j$. These functions are denoted by $v_\rho(nR'_j, p)$ where the suffix ρ accounts for the position of the cells and ranges from 1 to $n!/\Pi_j(nR'_j!)$ and where

$$W(nR'_j, p) = \sum_\rho v_\rho(nR'_j, p)$$

is the function of distribution of the configurations with assigned numbers R'_j.

Since the sets of configurations are specified by assigning to every cell a class of configurations, and since the x-chain may be regarded as an assembly of independent aggregates, equation (14) applies to the present problem The functions $v_\rho(nR'_j, p)$ approach accordingly independence of the suffix and proportionality to $W(nR'_j, p)$. The factor of proportionality remains undetermined and may depend on the numbers R'_j. Those functions $v_\rho(nR'_j, p)$ which deviate appreciably from the average are negligibly small or identically 0.

If alternatively the neighbouring x-chains have the co-ordinates y and $y-1$, similar sets of configurations of the first chain are defined by assigning to every cell one particular class $A(y)_i$. The function of distribution of the set approaches asymptotically proportionality to the function $W(nR_i, p)$, where nR_i is the number of cells to which the class $A(y)_i$ is assigned.

Finally, consider those sets of configurations of a chain in which a class $A(y)_i$ and a class $B(y)_j$, in other words a matrix element $D(y)_{ij}$, are assigned to every cell. In sets of this kind either all configurations or no configuration combine with any arbitrary pair of configurations of the two neighbouring x-chains. The corresponding function of distribution depends upon the numbers of cells nR^+_{ij} to which simultaneously the classes $A(y)_i$ and $B(y)_j$ are assigned and is—similarly as the functions $W(nR'_j, p)$ and $W(nR_i, p)$—independent of the position of the cells with assigned classes. It is denoted by $U(nR^+_{ij}, p)$.

6. The statistics of a layer and the lattice

In § 4 it is shown that the function of distribution of an xy-layer is equal to $\sum_{ij} L^{n_{ij}}$, where L is a symmetric matrix the elements of which contain all configurations of the x-chain.

The x-chains which are to be joined may be tentatively replaced by assemblies of independent cells which occupy the configurations c_λ with the probabilities p'_λ. For this problem the matrix $H(y)^{[n]}$ is appropriate where $H(y)$ is a matrix obtained from $G(y)$ by substituting p'_λ for p_λ and $H(y)^{[n]}$ is the nth *outer* power of $H(y)$. An xy-layer consisting of m x-chains should according to this tentative approach have the function of distribution $\sum_{i,j}[(H(y)^{[n]})^m]_{ij}$. Since outer multiplication commutes with matrix multiplication the latter expression is equal to $\left[\sum_{i,j} H(y)^m{}_{ij}\right]^n$ which is obviously the function of distribution of an assembly of n independent y-chains.

The matrix $H(y)^{[n]}$ may be partitioned into (square or rectangular) submatrices in each of which all configurations with assigned numbers R^+_{ij} are contained. All matrix elements within a submatrix are equal to each other and proportional to the function of distribution for the configurations of this submatrix which is obviously equal to $U(nR^+_{ij}, p)$.

The matrix L is defined in the same space as the matrix $H(y)^{[n]}$ and its structure is similar. In each submatrix the function of distribution is equal to $U(nR^+_{ij}, p)$; the matrix elements are no longer equal; the sums of subrows and subcolumns, however, are either equal to each other or negligibly small, possibly equal to 0.

The sums of those subrows and subcolumns which give an appreciable contribution to the function of distribution are proportional to $U(nR^+_{ij}, p)$. The factor of proportionality differs for the various submatrices, i.e. depends upon the numbers R^+_{ij}. For a given temperature it is sufficient to consider only a small range of these numbers where the factor of proportionality may be regarded as constant or rather as an unknown function of temperature.

In calculating the powers of L two results of matrix algebra are required. They are given without proof since they may readily be verified by means of the matrix law of multiplication.

(*a*) If a square matrix is partitioned into submatrices such that the partitioning lines are symmetric about the leading diagonal, the powers of this matrix may be calculated by regarding the submatrices as matrix elements. The product of two submatrices is defined as matrix product in the usual way.

(*b*) Two square or rectangular matrices for which multiplication is defined and in each of which the sum of elements is equal for each row and column have a product in which the sum of elements is equal for each row and column. If in either factor some rows and columns contain only elements equal to 0, while for the remaining rows and columns the sums of elements are equal, the product contains some rows and columns with all elements equal to 0 while for the remaining rows and columns the sums of elements are equal.

In applying these theorems to the powers of L it is seen that instead of multiplying a submatrix of L^m with a submatrix of L, the corresponding submatrices of $(H(y)^{[n]})^m$ and $H(y)^{[n]}$ may be multiplied. The resulting functions of distributions differ only by a factor which may depend upon t but not upon the p_λ.

The above tentative approach to the function of distribution of the xy-layer is therefore justified. The function of distribution of the xy-layer is equal to

$$\sum_{i,j} L^{mij} = \alpha(t) \cdot \left[\sum_{i,j} H(y)^{mij} \right]^n, \tag{19}$$

where $\alpha(t)$ is an unknown function of temperature.

Provided that $H(y)$ is a G-matrix the statistics of an xy-layer may be shown to be the statistics of an assembly of independent aggregates, similarly as it is shown for an x-chain. The arguments which lead from the function of distribution of the x-chain to the function of distribution of the xy-layer may be applied to the latter and lead to the function of distribution of the lattice It is sufficient to give the result.

The configurations c_λ are supposed to be contained in the matrix elements $D(x)_{gh}$, $D(y)_{ij}$, $D(z)_{kl}$. The mean probability coefficient of the configuration c_λ is denoted by p'_λ, p''_λ, p'''_λ when referring to a cell which is sited within a chain, a layer or the lattice respectively. $K(z)$ is a matrix obtained by substituting p''_λ for p_λ in the matrix $G(z)$.

A range of temperature is considered in which $G(x)$, $H(y)$ and $K(z)$ are G-matrices. The components of their first proper vectors are denoted by ξ_g, η_i, ζ_k respectively. It has been shown that

$$p'_\lambda = p_\lambda \xi_g \xi_h, \tag{18}$$

and it may similarly be proved that

$$p''_\lambda = p_\lambda \xi_g \xi_h \eta_i \eta_j, \quad p'''_\lambda = p_\lambda \xi_g \xi_h \eta_i \eta_j \zeta_k \zeta_l. \tag{20}$$

The function of distribution for a cell which is sited within the lattice is therefore equal to

$$P = \sum_\lambda t^{E_\lambda} \xi_g \xi_h \eta_i \eta_j \zeta_k \zeta_l. \tag{21}$$

P is not equal to the partition function since an unknown function of temperature enters into equation (19) and a similar function enters into the function of distribution of the lattice. In § 3 E_λ is defined as the energy of the configuration c_λ and $t^{E_\lambda} = \exp[-E_\lambda/kT]$. In evaluating equation (21), we consider the energies given as multiples of an arbitrary unit of energy which is denoted by ϵ. Then $t = \exp[-\epsilon/kT]$. The thermodynamic energy per cell and also per domain of the crystal is obtained by taking an average of the cell energy

$$\bar{E} = \epsilon t \, \partial \log P / \partial t \tag{22}$$

In carrying out the differentiation $\xi_g \xi_h \eta_i \eta_j \zeta_k \zeta_l$ is regarded as independent of t.

By means of equations (21) and (22) the thermodynamic energy can be calculated as function of temperature. If the theory is applied to a special system such as specified in the introduction, this relation between temperature and energy is capable of experimental verification.

References

Eisenschitz, R. 1941 *Nature, Lond.*, **147**, 778.
Lassettre, E. N. & Hove, J. P. 1941 *J. Chem. Phys.* **9**, 747, 808.
Montroll, E. W. 1941 *J. Chem. Phys.* **9**, 706.
Temple, G. 1934 *The general principles of quantum theory.* London.

Matrix theory of correlations in a lattice. Part II

By R. Eisenschitz

(*Communicated by Sir Robert Robertson, F.R.S.—Received* 5 *March* 1942—*Revised* 25 *May* 1943)

The specific-heat curve corresponding to order-disorder equilibrium is derived from the nearest neighbour model by means of the general theory of correlations in a lattice For a lattice in two dimensions the resulting thermodynamic energy is an algebraic expression; for a lattice in three dimensions the thermodynamic energy is found by numerical methods. The c_v curve has a broad maximum and has no resemblance whatever to the experimental c_p curve. The reasons for this discrepancy are discussed.

1. Introduction

In this Part II the matrix theory of correlations as given in the preceding paper is applied to the order disorder equilibrium such as is known to exist in β-brass.

The system is represented by the nearest neighbour model which is recalled briefly in the following. A binary alloy is considered which forms a simple quadratic or cubic lattice. The ratio of atoms of either kind is assumed to be variable; the resulting specific heat curve applies also to an equimolecular alloy. The energies of vibration and arrangement are assumed to be independent and the latter only is taken into account. Every configuration has equal statistical weight. The energy is assumed to be equal to the sum of bond energies between nearest neighbours. Every bond between atoms of the same kind contributes the positive energy ϕ, the other bonds contributing 0.

From this model the specific heat at constant volume* will be derived without introducing additional assumptions or dubious approximations. In a preliminary note (Eisenschitz 1941) the result has been given for the lattice in two dimensions. In this paper the calculation is carried out for the lattice in two and three dimensions.

2. The lattice in two dimensions

In order to apply the general theory an appropriate domain has to be chosen. A square of four lattice points which are nearest neighbours is taken as domain. The bonds between these lattice points are denoted by a, b, c, d (figure 1). Since every bond is shared by two domains, it contributes to the energy of the domain the energies 0 or $\frac{1}{2}\phi$ respectively.

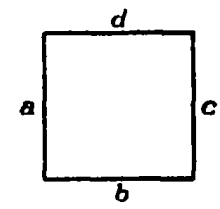


Figure 1

The configurations of the domain are found by distributing atoms of two kinds in all possible ways over the corners of the square. Sixteen different configurations are obtained in this way If, however, the configurations are specified in terms of bond energies, the number of configurations is reduced to eight. This specification i

* The theory is not concerned with the stability of the superlattice.

sufficient for calculating the thermodynamical energy. The eight configurations of the domain are shown in figure 2 in which a dot denotes the lattice point and a line denotes a bond, the energy of which is equal to $\frac{1}{2}\phi$.

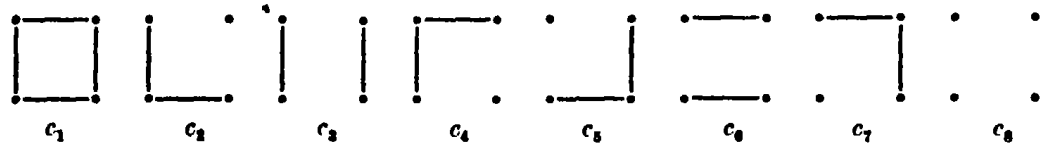


FIGURE 2

In figure 3 a configuration of a section of the lattice is shown. Every square corresponds to one domain. The two adjacent sides of neighbouring squares correspond to one and the same bond; if one of these sides is marked by a line, the other is necessarily also marked. A pattern of this kind is a picture of the fictitious lattice which is introduced in Part I, § 3; one square is the 'cell' of the fictitious lattice.

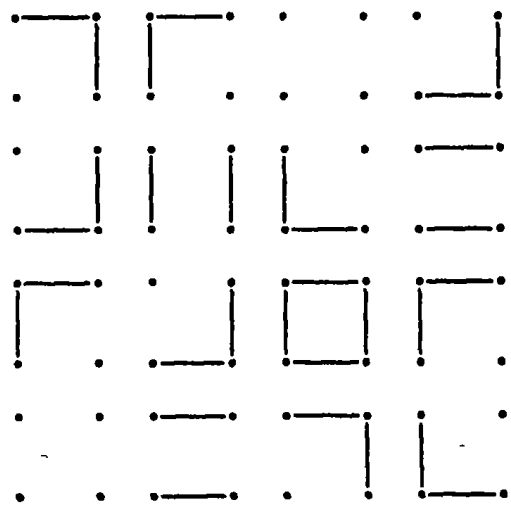


FIGURE 3

In order to specify the configurations of the cell without reference to the figure they are denoted by symbolic product of a, b, c, d to which every bond having the energy $\frac{1}{2}\phi$ contributes its symbol.

$$c_1 = abcd, \quad c_2 = ab, \quad c_3 = ac, \quad c_4 = ad,$$
$$c_5 = bc, \quad c_6 = bd, \quad c_7 = cd, \quad c_8 = 1.$$

The configurations are referred to as 'a-configurations' and 'non-a-configurations', etc., if there is or is not an 'a' in the symbolic product by which they are denoted. The energies of the configurations (expressed as multiples of ϕ) and their probability coefficients are

$$E_1 = 2, \quad E_2 = E_3 = E_4 = E_5 = E_6 = E_7 = 1, \quad E_8 = 0,$$
$$p_1 = t^2, \quad p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = t, \quad p_8 = 1,$$

where $t = \exp(-\phi/kT)$.

Two cells with the co-ordinates x and $x+1$ ('x-neighbours') are sharing one bond which is the c-bond of the first and the a-bond of the second cell. The second cell necessarily occupies an a-configuration when the first occupies a c-configuration, and the second occupies a non-a-configuration when the first occupies a non-c-configuration. Two cells with the co-ordinates y and $y+1$ are sharing one bond which is the d-bond of the first and the b-bond of the second cell. The latter occupies a b-configuration when the first occupies a d-configuration and the latter occupies a non-b-configuration when the first occupies a non-d-configuration. The probability coefficient of any configuration of a pair of neighbouring cells is obtained by multiplying the corresponding probability coefficients of the two cells.

These are the rules of composition, from which the configurations of the lattice and their probability coefficients are subsequently derived In the terminology as used in Part I, § 3 the configurations c_λ are in four different ways divided into classes:

$A(x)_1$: c_1, c_2, c_3, c_4	a-configurations
$A(x)_2$: c_5, c_6, c_7, c_8	non-a-configurations
$A(y)_1$: c_1, c_2, c_5, c_6	b-configurations
$A(y)_2$: c_3, c_4, c_7, c_8	non-b-configurations
$B(x)_1$ c_1, c_3, c_5, c_7	c-configurations
$B(x)_2$: c_2, c_4, c_6, c_8	non-c-configurations
$B(y)_1$ c_1, c_4, c_6, c_7	d-configurations
$B(y)_2$: c_2, c_3, c_5, c_8	non-d-configurations

The matrices $D(x)$, $D(y)$, $G(x)$, $G(y)$ are defined in Part I, § 3:

$$D(x) = \begin{Vmatrix} c_1+c_3 & c_2+c_4 \\ c_5+c_7 & c_6+c_8 \end{Vmatrix}, \qquad D(y) = \begin{Vmatrix} c_1+c_6 & c_5+c_2 \\ c_7+c_4 & c_3+c_8 \end{Vmatrix},$$

$$G(x) = G(y) = \begin{Vmatrix} t^2+t & 2t \\ 2t & t+1 \end{Vmatrix}.$$

The matrix $D(x)^n$ contains obviously all those power products of a, b, c, d which correspond to the possible configurations of an x-chain of n cells and the matrix $G(x)^n$ contains the corresponding probability coefficients

It is readily verified that $G(x)$ is a G-matrix as defined in Part I, § 3 Its proper values are denoted by G' and G'' and, writing G_{gh} for $G(x)_{gh}$, equal to

$$(1/2)\,[G_{22}+G_{11} \pm \sqrt{\{(G_{22}-G_{11})^2-4G_{12}^2\}}] = \tfrac{1}{2}[(1+t)^2 \pm \sqrt{(1+14t^2+t^4)}],$$

$G' > 0$ and $G' > G''$ and $G'^2 > G''^2$ for $0 < t \leqslant 1$. The components of the first proper vector are

$$\xi_1 = G_{12}/\sqrt{[G_{12}^2+(G'-G_{11})^2]}, \quad \xi_2 = (G'-G_{11})/\sqrt{[G_{12}^2+(G'-G_{11})^2]}.$$

So far the references to the general theory are readily verified from the model. In the following such verification would be possible only by an almost word by word repetition of the general theory. As examples for the main notions of the theory we

give expressions for the numbers R_{ij}^{+}, the mean probability coefficients p'_{λ} and the matrix $H(y)$. Only the latter enter into the result.

The configurations of the x-chain may be specified by the populations of the configurations $c_1 \dots c_8$ which are denoted by $nr_1 \dots nr_8$. $r_1+r_2+\dots+r_8 = 1$. The numbers R_{ij}^{+} which are introduced in Part I, § 5 are expressed in terms of the numbers r_{λ}.

$$R_{11}^{+} = r_1+r_6, \quad R_{12}^{+} = r_3+r_5,$$
$$R_{21}^{+} = r_4+r_7, \quad R_{22}^{+} = r_2+r_8.$$

The function of distribution of the chain configurations with assigned values of R_{11}^{+}, R_{12}^{+}, R_{21}^{+}, R_{22}^{+} is denoted by $U(nR_{ij}^{+}, p)$. In Part I, § 5, it is shown that this function is independent of the position of those nR_{11}^{+} cells which may occupy the configurations c_1 or c_6 and in a similar way independent of the positions of the other cells. On this fact the result is based which is given in Part I, equations (21) and (22). These equations are evaluated in the following.

According to Part I, equation (18), the mean probability coefficients for a cell that is sited within an x-chain are

$$p'_1 = t^2\xi_1^2, \quad p'_2 = p'_4 = p'_5 = p'_7 = t\xi_1\xi_2,$$
$$p'_3 = t\xi_1^2, \quad p'_6 = t\xi_2^2, \qquad p'_8 = \xi_2^2.$$

By substituting p'_{λ} for p_{λ} in the matrix $G(y)$, the matrix $H(y)$ is obtained:

$$H(y) = \begin{Vmatrix} t^2\xi_1^2+t\xi_2^2 & 2t\xi_1\xi_2 \\ 2t\xi_1\xi_2 & t\xi_1^2+\xi_2^2 \end{Vmatrix}.$$

By calculating the proper values of $H(y)$ it may be readily verified that it is a G-matrix, which fact is essential for the validity of the result.

The function of distribution for a cell that is sited within the lattice is equal to H', the higher proper value of the matrix $H(y)$ H' depends explicitly on t and upon ξ_1 and ξ_2 which are also functions of t By substituting in ξ_1 and ξ_2 a new variable θ for t we obtain H' as function of t and θ According to Part I, equation (22), and putting $\epsilon = \phi, \bar{E}$, the thermodynamic energy per cell (which is equal to the thermodynamic energy per atom), is found

$$\bar{E} = \phi \quad t(\partial \log H'/\partial t).$$

H' is calculated by solving the characteristic equation of the matrix $H(y)$ and expressing ξ_1 and ξ_2 as functions of θ

$$\begin{aligned} H' = (1/2)\{&(1+t)[(1-\theta^2)^2+8\theta^2(1+t)+(1-\theta^2)\sqrt{(1+14\theta^2+\theta^4)}] \\ &+\sqrt{2}[(1-\theta^2)^4(1-t)^2+32\theta^4(1+14t^2+t^4)+8\theta^2(1-\theta^2)^2(2-3t+8t^2+t^3) \\ &+\{(1-\theta^2)^3(1-t)^2+8\theta^2(1-\theta^2)(1-t+7t^2+t^3)\}\sqrt{(1+14\theta^2+\theta^4)}]^{\frac{1}{2}}\} \\ &\{(1+14\theta^2+\theta^4)+(1-\theta^2)\sqrt{(1+14\theta^2+\theta^4)}\}^{-1} \end{aligned}$$

The specific heat curve which follows from this equation is given in figure 5. It is discussed in § 4 together with the specific heat curve of the lattice in three dimensions.

3. The lattice in three dimensions

The cell of the fictitious lattice is a cube the bonds of which are denoted by $A \ldots h$ (figure 4). As every bond is shared by four cells, its energies per cell are 0 and $\frac{1}{4}\phi$. There are 128 configurations c_λ which are represented by symbolic products. The matrices $D(x)$, $D(y)$, $D(z)$, $G(x)$, $H(y)$, $K(z)$ are given in tables 1–4; in these tables $t = \exp(-\phi/4kT)$. The matrices comply with the conditions of symmetry stated in Part I, § 4. The number of configurations is equal in all their rows.

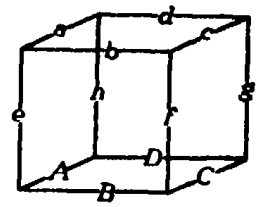


Figure 4

The function P is calculated in terms of the components of the first proper vectors. Denoting the product $\xi_g\xi_h$, $\eta_i\eta_j$, $\zeta_k\zeta_l$ by a bracket (gh, ij, kl), it is found that

$$\begin{aligned}P = {} & (88,88,88) + 8t^3(28,28,28) + t^4[(33,88,33) + (33,66,88) + (88,33,66) \\ & + 4(38,38,22) + 4(22,68,38) + 4(38,22,68)] + 8t^5[(23,28,23) \\ & + (23,26,28) + (28,23,26)] + 4t^6[(22,22,18) + (22,22,36) + (22,18,22) \\ & + (22,36,22) + (18,22,22) + (33,22,22) + 2(22,22,22)] + 8t^7[(23,23,12) \\ & + (23,12,26) + (12,26,23)] + t^8[(33,33,11) + (33,11,66) + (11,66,33) \\ & + 4(13,22,13) + 4(13,16,22) + 4(22,13,16)] + 8t^9(12,12,12) \\ & + t^{12}(11,11,11),\end{aligned}$$

and according to equation (22) of Part I with $\epsilon = \frac{1}{4}\phi$,

$$\bar{E} = (\phi t/4)\,\partial \log P/\partial t.$$

The components of the proper vectors are calculated by a method of approximation.

As the cell, the x-chain and the xy-layer have at least digonal symmetry round the x-, y-, z-axis respectively the matrices $G(x)$, $H(y)$, $K(z)$ are transformed by an orthogonal matrix into a system of basic vectors forming subspaces that are invariant towards digonal rotations By means of the matrix of transformation

$$U' = \left\| \begin{array}{cccccccc} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right\|,$$

and writing G_{ik} for $G(x)_{ik}$, five rows of $U'^{-1}GU$ are obtained:

$$\Gamma = \left\| \begin{array}{ccccc} G_{88} & G_{86} & G_{83} & G_{81} & 2G_{82} \\ G_{68} & G_{66} & G_{63} & G_{61} & 2G_{62} \\ G_{38} & G_{36} & G_{33} & G_{31} & 2G_{32} \\ G_{18} & G_{16} & G_{13} & G_{11} & 2G_{12} \\ 2G_{28} & 2G_{26} & 2G_{23} & 2G_{21} & 3G_{24} + G_{22} \end{array} \right\|.$$

The remaining diagonal elements are equal to $G_{22}-G_{24}$ while the non-diagonal elements are vanishing.

If $t \ll 1$, then Γ is approximately diagonal so that the required transformation is infinitesimal. The matrix of transformation is equal to

$$U'' = \begin{Vmatrix} 1/N_1 & -G_{84}/G_{88}N_2 & -G_{83}/G_{88}N_3 & -G_{81}/G_{88}N_4 & -2G_{82}/G_{88}N_5 \\ G_{84}/G_{88}N_1 & 1/N_2 & 0 & 0 & 0 \\ G_{83}/G_{88}N_1 & 0 & 1/N_3 & 0 & 0 \\ G_{81}/G_{88}N_1 & 0 & 0 & 1/N_4 & 0 \\ 2G_{82}/G_{88}N_1 & 0 & 0 & 0 & 1/N_5 \end{Vmatrix}$$

where $N_1 \ldots N_5$ are factors of normalization which do not enter into the final result. If $t = 1$, then $U''^{-1}\Gamma U''$ is diagonal. U'' is used accordingly for all values of t as

TABLE 1. $D(x)$

	abcd	*ab*	*ac*	*ad*	*bc*	*bd*	*cd*	1
ABCD	*efgh*	*efh*	*eh*	*egh*	*efg*	*ef*	*fgh*	*fh*
	+	+	+	+	+	+	+	+
	1	*g*	*fg*	*f*	*h*	*gh*	*e*	*eg*
AB	*efh*	*efgh*	*egh*	*eh*	*ef*	*efg*	*fh*	*fgh*
	+	+	+	+	+	+	+	+
	g	1	*f*	*fg*	*gh*	*h*	*eg*	*e*
AC	*eh*	*egh*	*efgh*	*efh*	*fgh*	*fh*	*efg*	*ef*
	+	+	+	+	+	+	+	+
	fg	*f*	1	*g*	*e*	*eg*	*h*	*gh*
AD	*egh*	*eh*	*efh*	*efgh*	*fh*	*fgh*	*ef*	*efg*
	+	+	+	+	+	+	+	+
	f	*fg*	*g*	1	*eg*	*e*	*gh*	*h*
BC	*efg*	*ef*	*fgh*	*fh*	*efgh*	*efh*	*eh*	*egh*
	+	+	+	+	+	+	+	+
	h	*gh*	*e*	*eg*	1	*g*	*fg*	*f*
BD	*ef*	*efg*	*fh*	*fgh*	*efh*	*efgh*	*egh*	*eh*
	+	+	+	+	+	+	+	+
	gh	*h*	*eg*	*e*	*g*	1	*f*	*fg*
CD	*fgh*	*fh*	*efg*	*ef*	*eh*	*egh*	*efgh*	*efh*
	+	+	+	+	+	+	+	+
	e	*eg*	*h*	*gh*	*fg*	*f*	1	*g*
1	*fh*	*fgh*	*ef*	*efg*	*egh*	*eh*	*efh*	*efgh*
	+	+	+	+	+	+	+	+
	eg	*e*	*gh*	*h*	*f*	*fg*	*g*	1

$D(y)$ and $D(z)$ are obtained from $D(x)$ by means of the permutations:

a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*A*	*B*	*C*	*D*	and	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*A*	*B*	*C*	*D*
e	*B*	*f*	*b*	*A*	*C*	*c*	*a*	*h*	*D*	*g*	*d*		*A*	*e*	*a*	*h*	*B*	*b*	*d*	*D*	*C*	*f*	*c*	*g*

respectively.

approximate matrix of transformation. Similarly the transformation of $H(y)$ and $K(z)$ is carried out. Omitting factors of normalization, the components of the proper vectors are found to be

$$\begin{array}{lll}
\xi_1 = 2t^6 & \eta_1 = \xi_1\xi_2^2 & \zeta_1 = \eta_1\eta_2^2 \\
\xi_2 = t^3+t^5 & \eta_2 = t^3\xi_2\xi_8+t^5\xi_2\xi_3 & \zeta_2 = t^3\xi_2\xi_8\,\eta_2\eta_8+t^5\xi_2\xi_3\,\eta_2\eta_6 \\
\xi_3 = 2t^4 & \eta_3 = \xi_3^2\xi_8 & \zeta_3 = \eta_6^2\eta_8 \\
\xi_6 = \xi_3 & \eta_6 = \xi_2^2\xi_3 & \zeta_6 = \eta_2^2\eta_3 \\
\xi_8 = 1+t^4 & \eta_8 = \xi_8^2+t^4\xi_3^2 & \zeta_8 = \xi_8^2\eta_8^2+t^4\xi_3^2\eta_6^2 \\
\xi_4 = \xi_5 = \xi_7 = \xi_2 & \eta_4 = \eta_5 = \eta_7 = \eta_2 & \zeta_4 = \zeta_5 = \zeta_7 = \zeta_2
\end{array}$$

TABLE 2. $G(x)$

$t^{12}+t^6$ t^9+t^7	t^9+t^7 t^8+t^4	$2t^6$ t^7+t^5	t^9+t^7 $2t^6$	t^9+t^7 $2t^6$	$2t^6$ t^7+t^5	t^9+t^7 $2t^6$	$2t^6$ t^5+t^3
$2t^6$ t^9+t^7	t^7+t^5 $2t^6$	t^6+t^4 t^7+t^5	t^7+t^5 t^6+t^4	t^7+t^5 $2t^6$	$2t^6$ t^7+t^5	t^7+t^5 $2t^6$	$2t^4$ t^5+t^3
t^9+t^7 $2t^6$	$2t^6$ t^7+t^5	t^7+t^5 $2t^6$	$2t^6$ t^7+t^5	t^6+t^4 t^7+t^5	t^7+t^5 $2t^6$	$2t^6$ t^5+t^3	t^5+t^3 $2t^4$
t^9+t^7 $2t^6$	$2t^6$ t^5+t^3	t^7+t^5 $2t^4$	$2t^6$ t^5+t^3	t^7+t^5 t^5+t^3	t^6+t^4 $2t^4$	t^7+t^5 t^5+t^3	$2t^4$ t^4+1

TABLE 3. $H(y)$

$(gh) = \xi_g\xi_h$

t^{12} (11) + t^6 (33)	t^9 (12) + t^7 (23)	t^8 (22) + t^6 (22)	t^9 (12) + t^7 (23)	t^9 (12) + t^7 (23)	t^8 (13) + t^5 (13)	t^9 (12) + t^7 (23)	t^6 (22) + t^6 (22)
t^9 (12) + t^7 (23)	t^8 (13) + t^4 (38)	t^7 (23) + t^5 (28)	t^6 (22) + t^6 (22)	t^6 (18) + t^6 (33)	t^7 (12) + t^5 (23)	t^6 (22) + t^6 (22)	t^5 (23) + t^3 (28)
t^8 (22) + t^6 (22)	t^7 (23) + t^5 (28)	t^6 (33) + t^4 (88)	t^7 (23) + t^5 (28)	t^7 (23) + t^5 (28)	t^6 (22) + t^6 (22)	t^7 (23) + t^5 (28)	t^4 (38) + t^4 (38)
t^9 (12) + t^7 (23)	t^6 (22) + t^6 (22)	t^7 (23) + t^5 (28)	t^8 (13) + t^4 (38)	t^6 (22) + t^6 (22)	t^7 (12) + t^5 (23)	t^6 (33) + t^6 (18)	t^5 (23) + t^3 (28)
t^9 (12) + t^7 (23)	t^6 (18) + t^6 (33)	t^7 (23) + t^5 (28)	t^6 (22) + t^6 (22)	t^5 (13) + t^4 (38)	t^7 (12) + t^5 (23)	t^6 (22) + t^6 (22)	t^5 (23) + t^3 (28)
t^5 (13) + t^6 (13)	t^7 (12) + t^5 (23)	t^6 (22) + t^6 (22)	t^7 (12) + t^5 (23)	t^7 (12) + t^5 (23)	t^6 (11) + t^4 (33)	t^7 (12) + t^5 (23)	t^4 (22) + t^4 (22)
t^9 (12) + t^7 (23)	t^6 (22) + t^6 (22)	t^7 (23) + t^5 (28)	t^6 (33) + t^6 (18)	t^6 (22) + t^6 (22)	t^7 (12) + t^5 (23)	t^5 (13) + t^4 (38)	t^5 (23) + t^3 (28)
t^6 (22) + t^6 (22)	t^5 (23) + t^3 (28)	t^4 (38) + t^4 (38)	t^5 (23) + t^3 (28)	t^5 (23) + t^3 (28)	t^4 (22) + t^4 (22)	t^5 (23) + t^3 (28)	t^4 (33) + (88)

By numerical calculation it was checked that $G(x)$, $H(y)$ and $K(z)$ are G-matrices.

As example for the error due to the method of approximation the transformation of $H(y)$ is given for $t = 0{\cdot}83527$, the temperature where the specific heat has its maximum.

The symmetric matrix

2·6717	0·95281	1·4090	0·66477	2·4974
	0·57059	0·66477	3·1332	1·1643
		1·3005	0·46379	1·7424
			0·277742	0·81232
				2·8555

is transformed on the approximately diagonal form

6·8715	0·27326	0 44954	0·18605	0·4089
	0·2050	0·1352	0·0679	0·1883
		0 4361	0·0972	0·2748
			0 1058	0·1354
				0·2780

TABLE 4. $K(z)$

$(gh, ij) = \xi_g \xi_h \eta_i \eta_j$

$t^{12}(11,11)$ + $t^{8}(33,33)$	$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{8}(13,22)$ + $t^{6}(13,22)$	$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{8}(22,13)$ + $t^{6}(22,13)$	$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{6}(22,22)$ + $t^{6}(22,22)$
$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{8}(13,16)$ + $t^{4}(38,38)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{8}(18,22)$ + $t^{6}(33,22)$	$t^{8}(22,18)$ + $t^{6}(22,36)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{8}(22,22)$ + $t^{6}(22,22)$	$t^{5}(23,26)$ + $t^{3}(28,28)$
$t^{8}(13,22)$ + $t^{6}(13,22)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{8}(11,66)$ + $t^{4}(33,88)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{6}(22,22)$ + $t^{6}(22,22)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{4}(22,68)$ + $t^{4}(22,68)$
$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{6}(18,22)$ + $t^{6}(33,22)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{8}(13,16)$ + $t^{4}(38,38)$	$t^{6}(22,22)$ + $t^{6}(22,22)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{6}(22,36)$ + $t^{6}(22,18)$	$t^{5}(23,26)$ + $t^{3}(28,28)$
$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{6}(22,18)$ + $t^{6}(22,36)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{6}(22,22)$ + $t^{6}(22,22)$	$t^{8}(13,16)$ + $t^{4}(38,38)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{6}(33,22)$ + $t^{6}(18,22)$	$t^{5}(23,26)$ + $t^{3}(28,28)$
$t^{8}(22,13)$ + $t^{6}(22,13)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{6}(22,22)$ + $t^{6}(22,22)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{8}(33,11)$ + $t^{4}(88,33)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{4}(38,22)$ + $t^{4}(38,22)$
$t^{9}(12,12)$ + $t^{7}(23,23)$	$t^{6}(22,22)$ + $t^{6}(22,22)$	$t^{7}(12,26)$ + $t^{5}(23,28)$	$t^{6}(22,36)$ + $t^{6}(22,18)$	$t^{6}(33,22)$ + $t^{6}(18,22)$	$t^{7}(23,12)$ + $t^{5}(28,23)$	$t^{8}(13,16)$ + $t^{4}(38,38)$	$t^{5}(23,26)$ + $t^{3}(28,28)$
$t^{6}(22,22)$ + $t^{6}(22,22)$	$t^{5}(23,26)$ + $t^{3}(28,28)$	$t^{4}(22,68)$ + $t^{4}(22,68)$	$t^{5}(23,26)$ + $t^{3}(28,28)$	$t^{5}(23,26)$ + $t^{3}(28,28)$	$t^{4}(38,22)$ + $t^{4}(38,22)$	$t^{5}(23,23)$ + $t^{3}(28,28)$	$t^{4}(33,66)$ + (88,88)

The ratio of the largest non-diagonal elements to the difference of the two highest diagonal elements is about 7 %. This shows that the two highest proper values are distinct and that $H(y)$ is a G-matrix.

4. Discussion

I have previously (1938) investigated the effect of thermal expansion on the specific heat of β-brass. It was found that small volume changes appreciably affect the specific heat. The theory of Bragg & Williams (1934) was shown to agree with experiments if the volume change is taken into account.

Special attention was given to the specific heat immediately beyond the λ-point. It was shown generally that the nearest neighbour model cannot account for any singularity of the specific heat at constant volume. Since the theory of Bragg & Williams gives a discontinuity of the specific heat, I tentatively introduced a modification of the theory. This part of the paper was criticized by Bethe & Kirkwood (1939). Neither my modification of the theory of Bragg & Williams nor the arguments of Bethe & Kirkwood can claim to be the correct consequence of the nearest neighbour model.

The correct specific heat at constant volume is obtained in this paper in §§ 2 and 3 and given in figure 5 as function of temperature; in this figure the curves 1, 2, 3 refer to the chain in one dimension and the lattice in 2 and 3 dimensions respectively These curves show no similarity to the λ-form of the experimental c_p curve.

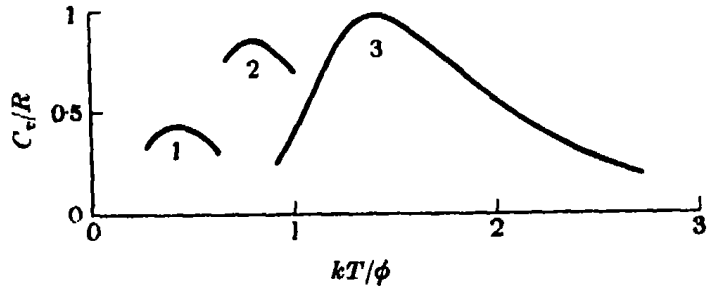


Figure 5

This discrepancy may be due to the effect of thermal expansion or, alternately, may indicate that the nearest neighbour model is inaccurate.

Good (1941) has measured the elastic constants of β-brass near the λ-point. By using his compressibility and the experimental coefficient of thermal expansion a rough estimate of $c_p - c_v$ has been made; the difference appears to be too small to account for the discrepancy. Good states, however, that the accuracy of his compressibility is not very great. A marked difference between the c_p and c_v curve has been found for NH_4Cl at the λ-point by Lawson (1940). His c_v curve has a flat and broad maximum.

The theory of the c_p curve involves two functions which have to be introduced in addition to the general assumptions of the nearest neighbour model: the bond energy

of atoms of the same kind and of atoms of different kinds as function of volume. Without making artificial assumptions on these energy curves it can be shown that the specific heat may be very sensitive to volume changes; a broad maximum of the c_v curve is, however, transformed into a broad maximum of the c_p curve which shows no similarity to the experimental λ-curves.

If it is assumed that the two energy curves intersect at a certain volume the specific heat at constant pressure has a λ-point even if the c_v curve has a flat maximum. This assumption seems to account for the discrepancy between theoretical and experimental specific heat; the experimental data as available are not sufficient to check it.

I am greatly indebted to the managers of the Royal Institution and to the late Sir William Bragg for the opportunity given for carrying out this work at the Davy Faraday Laboratory. I also wish to thank Dr A. Müller for his help and interest.

References

Bethe, H. A. & Kirkwood, J. G. 1939 *J Chem. Phys* **7**, 578.
Bragg & Williams. 1934 *Proc. Roy. Soc* A, **145**, 699.
Eisenschitz, R. 1938 *Proc Roy. Soc.* A, **168**, 546; *Nature, Lond.* (1941), **147**, 778.
Good, W. A. 1941 *Phys. Rev.* **60**, 605.
Lawson, A. W. 1940 *Phys. Rev.* **57**, 417.

The heat of adsorption of long-chain compounds and their effect on boundary lubrication

By Joseph John Frewing

(*Communicated by E. K. Rideal, F.R.S.—Received 27 January* 1943)

Shell Refining and Marketing Company, Limited, Britannia Laboratory, Cambridge

The frictional behaviour between mild steel surfaces lubricated with solutions in white oil of long-chain halides, acids, α-substituted acids, esters, cyanide, thiocyanate and a nitro derivative has been investigated under high loads at low speeds In all cases a transition from smooth sliding to stick-slips occurs at a temperature characteristic of the particular solution employed For each substance the transition temperature increases with the concentration.

Each solution builds up, and is in equilibrium with, an adsorbed and oriented film of the polar compound on the surface. Assuming that the transition occurs when the surface concentration of this film decreases to a certain value which, for any one material is independent of temperature, an equation has been deduced relating the concentration, and transition temperature with the heat of adsorption U. All the experimental results are in good agreement with this equation.

The values of U show that these long-chain polar compounds are adsorbed by the interaction of their dipoles with the atoms in the metal surface, and not by any chemical reaction.

The results also suggest that the esters are similarly oriented at metal and at aqueous surfaces.

An investigation of the effect of temperature on the boundary lubrication of mild steel surfaces, by a number of pure long-chain compounds, was recorded in a recent communication (Frewing 1942). The apparatus, which has already been fully described, consisted essentially of a hemispherical contact attached to a system of stiff restraint and low moment of inertia, for recording rapid fluctuations in friction The load was applied on the contact which rested on a flat plate driven at a uniform speed of *ca.* 0·005 cm./sec., and lubricated with the material under examination. It was shown that when a polar compound was used to lubricate such a system, the motion was at first smooth, but changed, on heating the surface, to the irregular stick and slip motion characteristic of liquid non-polar lubricants. The temperature at which this transition occurred was found, within the limit of experimental error, to be characteristic of the lubricant employed. The results obtained both for an excess of the polar lubricant, and for monolayers and multilayers, built up by the Langmuir-Blodgett technique, indicated that the transition from smooth sliding to stick-slips occurred when the adsorbed film became disoriented.

The investigation has now been extended to solutions of long-chain polar bodies in white oil,* a non-polar solvent. Each solution is found to have its characteristic transition temperature, and for each substance the transition temperature increases with the concentration.

* A highly refined, practically colourless, mineral oil. Specific gravity at 60/60° F, 0·888. Kinematic viscosities at 70 and 140° F, 86·5 and 14 centistokes respectively.

It is possible to explain these results by assuming that the polar body in the solution builds up, and is in equilibrium with, an adsorbed and oriented film on the surface. Smooth sliding occurs when a large proportion of the surface is covered with this film, but as the temperature rises the surface concentration of the adsorbed film decreases and the motion changes to stick-slips. If the transition occurs when the surface concentration decreases to a definite value which, for any one material, is independent of temperature, it is possible to deduce an equation relating the concentration in the solution with the transition temperature and the heat of adsorption. This equation is shown to hold for a large number of different polar compounds. The values for the heat of adsorption have thrown some light on the mechanism by which the molecules are adsorbed, and on their orientation at the surface.

Theoretical

Consider an adsorbed film on the surface in equilibrium with a solution containing the polar body, of which the film is composed, at a concentration C.

Let a fraction x of the surface be covered with this adsorbed film.

Now the molecules of the polar body will build up the adsorbed film at a rate proportional to the concentration C, and to the fraction of the surface not occupied by adsorbed molecules, i e. $(1-x)$.

The rate at which the film builds up is therefore

$$k_1 C(1-x),$$

when k_1 is constant at constant temperature.

The molecules adsorbed on the surface will leave the surface and pass into the solution at a rate proportional to the fraction of the surface which they occupy.

The rate at which the film breaks down will therefore be

$$k_2 x,$$

when k_2 is constant at constant temperature.

When the film is in equilibrium with the solution the rate at which the film is being formed must equal the rate at which it is disintegrating.

Therefore, at equilibrium

$$\frac{k_1}{k_2} = C\frac{x}{(1-x)} = K,$$

when K is the equilibrium constant.

The variation of K with temperature is given by the Van 't Hoff isochore

$$\frac{d \log_e K}{dT} = -\frac{U}{RT^2},$$

which on integration gives

$$\log_e K = \frac{U}{RT} + \text{integration constant},$$

where U is the heat of adsorption of the substance, from its state in the solution, on to the surface, R the gas constant and T the absolute temperature. This integration assumes that U is independent of T, which is probably justifiable over the small range of temperature investigated.

Assuming that the transition from smooth sliding to stick-slips occurs when the surface concentration of the adsorbed and oriented film decreases to a definite value—i.e. at $x=a$—which for any one material is independent of temperature, we have

$$\log_e \frac{a}{C(1-a)} = \frac{U}{RT_t} + \text{integration constant},$$

whence

$$2{\cdot}3 \log_{10} C = -\frac{U}{RT_t} + \text{constant},$$

where T_t is the value of the transition temperature on the absolute scale.

A straight line should therefore be obtained on plotting $\log_{10} C$ against $1/T_t$, the slope of this line being $-U/2{\cdot}3R$.

Experimental

The apparatus, technique and the mild steel contact and plate were the same as those used in the previous investigation

Since $\log_{10} C$ is plotted against $1/T_t$ in testing the theory developed above, any quantity which is proportional to the number of molecules in unit volume of the solution may be used for C, without affecting the slope of the $\log_{10} C - 1/T_t$ line. In all the following results the concentration is expressed as % wt. of solution. Since the densities of the long-chain compounds are very similar to that of white oil, and no appreciable change in volume occurs on solution, the error introduced in the slope of the $\log_{10} C - 1/T_t$ line through using % wt. instead of wt. per unit volume is negligible.

Some of the more concentrated solutions deposited crystals at room temperature, but in all cases the solutions were stable at temperatures below their transition temperatures so that it was possible to determine these values by taking care not to allow the plate to cool sufficiently for crystallization to occur.

The values of the transition temperature of each solution varied somewhat from one point of the surface to another, due to the difficulty of obtaining a surface of uniform activity. Each result given represents the mean of between ten and twenty determinations during the course of which the surface was cleaned and relapped several times.

In all cases the transition was reversible with temperature and, within the limit of experimental error, was independent of the number of times the solution was heated through the transition temperature. This shows that none of these polar compounds attacked the surface chemically.

The effect of load

When an excess of oleic acid was used as lubricant it was found that the coefficient of friction, in the region of smooth sliding below the transition temperature, increased with the load from 0·09 at ½ kg. reaching a limiting value of 0·16 at *ca.* 3 kg. When the surfaces were coated with a monolayer of, for example, ethyl stearate, the coefficient of friction remained constant at *ca.* 0·18 over the range of loads of ½–4 kg. This suggests that in the case of oleic acid orientation extends beyond the primary layer, and that the outer layers are able to carry some of the load. If the total load is sufficiently light an appreciable fraction of it is carried by these outer layers. With this apparatus it is necessary to use loads of 4 kg. or over in order to reduce, to negligible proportions, the load supported by the outer layers (Frewing 1942). Further support for the view that the high limiting value of the coefficient of friction is due to this cause, and not to the contact ploughing a track in the plate, is afforded by an experiment in which a second run was made over the same track, using pure oleic acid as lubricant. This resulted in a slight, but definite, increase in the coefficient of friction due presumably to the roughening of the surface by the first run. The 'ploughing term' proposed by Bowden & Tabor (1942) is therefore negligible under these conditions.

In order to ensure that conditions were those of true boundary lubrication, loads of 4–4½ kg. were used in all the experiments. The values of the coefficient of friction in the region of smooth sliding below the transition temperature usually varied between 0·15 and 0·22 for each solution, the mean value in each case being *ca.* 0·18. This agrees with that previously obtained with monolayers of polar compounds of the same chain length, and shows that conditions of true boundary lubrication were attained in these experiments with white oil solutions.

Some experiments were made on solutions of octadecyl chloride using loads of 8 kg. to see whether the transition now occurred at a different value of 'a', in which case the $\log_{10} C - 1/T_t$ line would have been displaced. The values obtained for the transition temperature are plotted against the concentration in figure 1. It will be seen that the points for both loads lie on the same curve. In figure 2 the values of $\log_{10} C$ are plotted against $1/T_t$. The values both for loads of 4 and 8 kg lie on the same straight line. It might have been expected that doubling the load would displace the $\log_{10} C - 1/T_t$ line, a higher surface concentration of adsorbed and oriented molecules now being necessary to give smooth sliding—in other words, the transition would occur at a higher value of 'a'. The hemispherical contact, however, will be plastically deformed under the load, the area of contact being proportional to the load (Bowden & Tabor 1939). The average pressure over the area of contact will therefore be independent of the load. The transition would therefore occur at the same value of 'a' irrespective of the load. This view is supported by the experimental result that the transition temperature is the same both with loads of 4 and 8 kg. over the range of concentration from 2 to 80 %. Increasing the load, however, will affect the response of the friction recorder, and it is possible

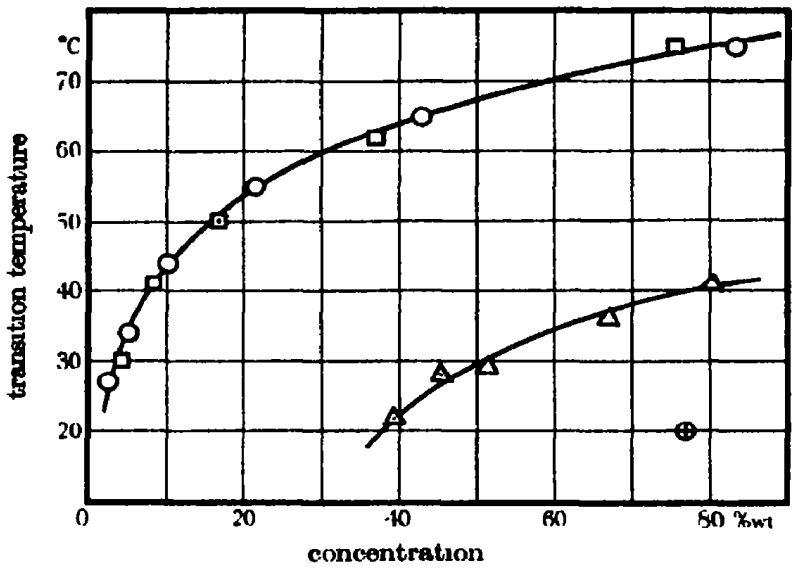


FIGURE 1. ⊙ octadecyl chloride, 4 kg.; □ octadecyl chloride, 8 kg., △ octadecyl iodide, 4 kg., ⊕ octadecyl bromide, 4 kg.

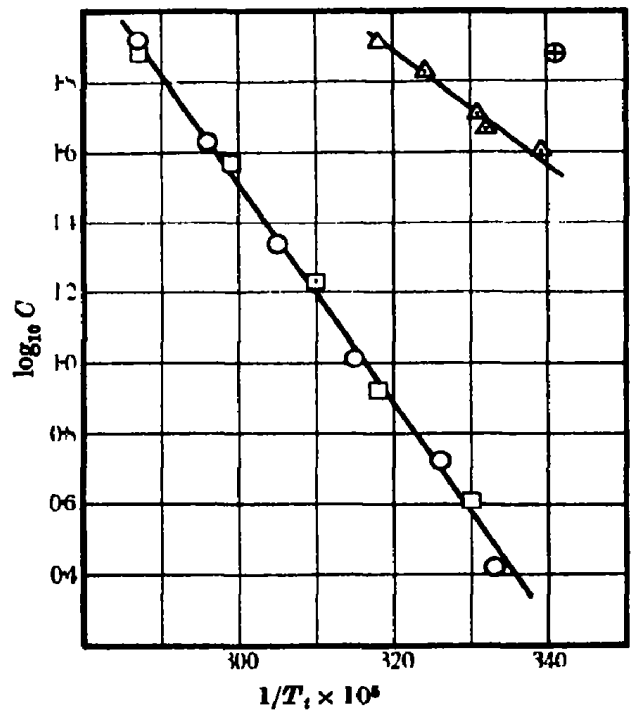


FIGURE 2. ⊙ octadecyl chloride, 4 kg.; □ octadecyl chloride, 8 kg.; △ octadecyl iodide, 4 kg., ⊕ octadecyl bromide, 4 kg.

that this may influence the transition temperature. Lower values for the transition temperature of oleic acid were obtained when a system of lower frequency was employed, but the low values of the coefficient of friction below the transition temperature suggested that the conditions were not those of true boundary lubrication (Frewing 1942). Hughes & Whittingham (1942), who used a smaller apparatus of the type described by Bowden & Leben (1939), with a steel contact on a cast iron plate, have also reported lower values for the transition temperatures of the fatty acids.

Octadecyl halides

The transition temperature—concentration curves for solutions of octadecyl chloride and octadecyl iodide are given in figure 1. The curve for solutions of octadecyl bromide might be expected to lie between those of the chloride and iodide, but a 77 % solution of this substance gave a transition temperature of only 20° C. It was therefore not possible to determine the curve for the bromide.

The values of $\log_{10} C$ plotted against $1/T_t$ are given in figure 2. Straight lines are obtained for both compounds. In the case of the chloride, where the concentrations investigated vary by a factor of 40, the agreement with the theory is particularly satisfactory. The values of U, the heat of adsorption of the molecules from their state in the solution, on to the mild steel surface, determined from the slopes of the lines are

octadecyl chloride, 14,500 cal /g. mol., octadecyl iodide, 7,500 cal /g. mol.

The value for the iodide must be regarded as approximate only, since it was not possible to examine this substance over a wide range of concentrations.

If the adsorption of the halides on to a steel surface is due to the interaction between the dipole of the carbon-halogen linkage and the surface, we should expect the chloride to be more strongly adsorbed than the iodide owing to the greater magnitude of the C-Cl dipole. The experimental values of the heat of adsorption support this view.

Fatty acids

The $\log_{10} C - 1/T_t$ lines for capric, myristic, stearic and oleic acids are shown in figure 3.

The values of U determined from the slopes of the lines in figure 3 are

capric acid, 12,500 cal /g. mol , stearic acid, 13,000 cal./g. mol.
myristic acid, 13,000 cal./g mol., oleic acid, 13,500 cal./g. mol.

These values are, within the limits of experimental error, identical. The value of U is therefore determined by the nature of the polar end group, and is not appreciably affected by the length of the hydrocarbon chain or whether—as in the case of oleic acid—a double bond is present at some distance from the end group.

It is interesting to note that the concentration of stearic, myristic and capric acids required to give a common transition temperature are in the approximate

proportion of 1:4.40 respectively, while the hydrocarbon chains of these acids contain 17, 13 and 9 carbon atoms respectively Now the experimental evidence indicates that, in the region of smooth sliding, the surfaces are covered with oriented adsorbed monolayers of the long-chain polar bodies which enable the contact to slide over the surface at a steady value of the friction. When the surface concentration of this monolayer decreases to a certain value, the contact is able to stick to the surface and travel forward with it until the limiting value of the static friction is reached. A rapid slip then occurs under the restoring force of the recording system, and the process is repeated, stick-slips occurring. If the acids are assumed

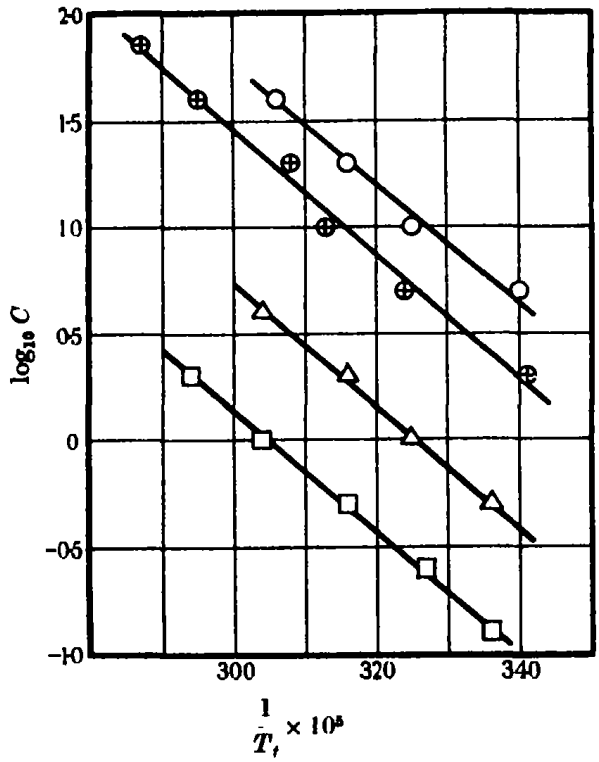


FIGURE 3. ⊙ capric acid; △ myristic acid, □ stearic acid; ⊕ oleic acid.

to be oriented vertically or at an approximately constant angle to the surface, it is possible that, on the average, the shorter the hydrocarbon chain the more closely the surfaces will be able to approach each other the attraction between them will consequently be greater. It is possible, therefore, that a higher surface concentration of the shorter adsorbed and oriented molecules would be required to support the load, in which case the transition would occur at a higher value of 'a' when the lower fatty acids are employed. The concentration of a fatty acid required to give a certain transition temperature would therefore decrease with increasing chain length.

As a result of hydrogen bonding the fatty acids exist, in the crystal lattice and in solution in non-polar solvents, as double molecules. The value of U determined

from the slope of the $\log_{10} C - 1/T_t$ line will therefore be the heat evolved when 1 g.mol. of the double molecules is adsorbed giving 2 g.mol. of adsorbed single molecules:

$$M_2 \rightarrow 2M \text{ (adsorbed)} + U.$$

This process may be represented by two stages·

$$M_2 \rightarrow 2M \text{ (solution)} - A,$$

where A is the heat of association, and

$$2M \text{ (solution)} \rightarrow 2M \text{ (adsorbed)} + 2Q,$$

where Q is the heat of adsorption of the single molecules.

Hence it follows that

$$Q = \tfrac{1}{2}(U + A)$$

Taking the heat of association of the fatty acids as *ca* 14,000 cal./g.mol. (see Herman 1940, for a review of the data), the heat of adsorption of the single molecules of the fatty acids therefore becomes *ca*. 13,500 cal./g.mol.

α-substituted fatty acids

It is well known that the introduction of a halogen atom into the α-position in a fatty acid increases the strength of the acid in aqueous solution. It is possible that the tendency for the proton to split off from the α-substituted acid is greater than in the case of the unsubstituted acid, even in non-aqueous media. If the adsorbed film were formed by some chemical union between the acid and the metal, it might therefore be supposed that the α-halogenated acid would form an adsorbed layer more readily than the simple acid. If the adsorbed film did result from a chemical reaction we should not expect the transition temperature to be independent of the number of times the surface was heated and cooled in contact with the solution. The heat of adsorption would also be higher for chemi-sorption than for dipole adsorption.

Solutions of α-bromo stearic acid and α-iodo stearic acid, however, gave their characteristic transition temperatures irrespective of the number of times they were heated in contact with the surface. Unfortunately it was not possible to prepare a pure specimen of α-chloro stearic acid, but α-hydroxy stearic acid which was made in the course of an attempt to obtain the α-chloro derivative was examined.

The results for these three α-substituted acids are given in figure 4. The results for stearic acid are included in figure 4 for comparison. The results for α-hydroxy stearic acid show excellent agreement with the theory down to the low concentration of 0·09 %.

The values of U for the α-substituted acids are:

α-bromo stearic acid, 10,000 cal./g.mol.; α-iodo stearic acid, 10,000 cal./g.mol.; α-hydroxy stearic acid, 13,500 cal./g.mol.

The effect of these substituents in the α-position on the degree of association in white oil and on the heat of association is, unfortunately, not known, but it is evident that taking these factors into account the heat of adsorption of the α-halogenated acids will not be very different from that of the fatty acids. This further supports the view that the adsorption is due to dipole interaction and not to a chemical process. Owing to the possibility of free rotation about single bonds, and the uncertainty of the orientation of the carboxyl group with respect to the surface it is not possible to estimate the effect of the introduction of another dipole on the resultant dipole moment resolved perpendicular to the surface.

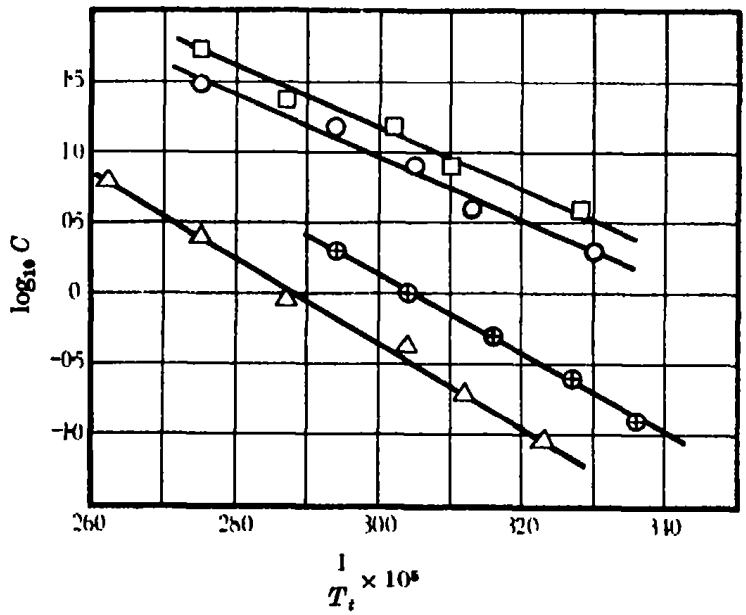


FIGURE 4 ⊙ α-bromo stearic acid; □ α-iodo stearic acid; ∧ α-hydroxy stearic acid, ∪ stearic acid.

Stearic esters of normal alcohols

The results obtained with solutions of methyl, ethyl, *n*-butyl, *n*-hexyl and *n*-octadecyl stearates are given in figure 5.

The values of U obtained from the slopes of the lines in figure 5 are:

methyl stearate, 8600 cal./g.mol., ethyl stearate, 3900 cal./g.mol.; *n*-butyl stearate, 3200 cal./g.mol.; *n*-hexyl stearate, 6900 cal./g.mol.; *n*-octadecyl stearate, 4900 cal./g.mol.

Malkin (1931) has shown by X-ray measurements that in the crystal lattice methyl stearate exists as double molecules, but ethyl stearate is present as single molecules. It is probable, therefore, that methyl stearate is associated in white oil, and that the esters of ethyl and higher alcohols are present in white oil solutions as single molecules. With the exception of methyl stearate, the above values of U may be taken as the heat of adsorption of single molecules on to the steel surface.

No information is available concerning the heat of association of the methyl esters, but this value is probably less than that of the acids so that the heat of adsorption of single molecules of methyl stearate will not be very different from the value of U given above.

It is interesting to note the effect of increasing the 'alcohol' chain on the value of the heat of adsorption This is highest in the case of the methyl esters, and decreases to a minimum when the 'alcohol' chain contains four carbon atoms; it increases with further increases in the length of the 'alcohol' chain to six carbon atoms, and then gradually decreases again as the chain is further lengthened.

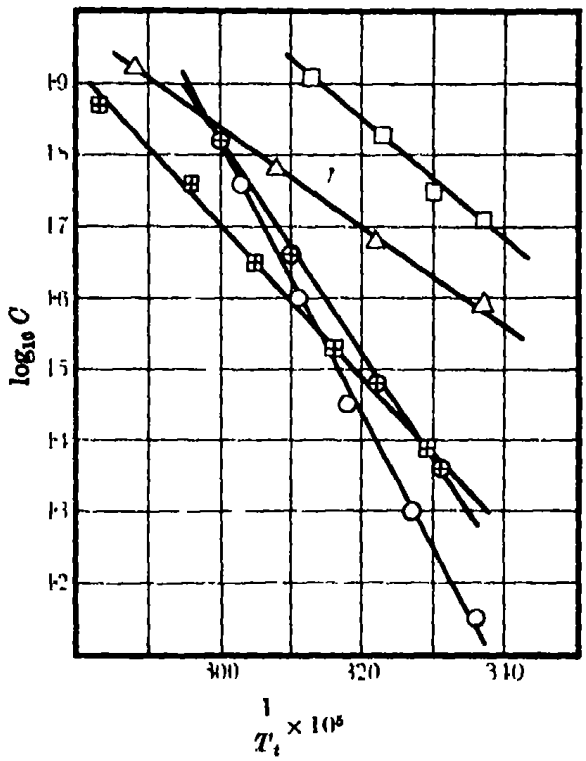


FIGURE 5. ○ methyl stearate, □ ethyl stearate; △ *n*-butyl stearate; ⊕ *n*-hexyl stearate; ⊞ *n*-octadecyl stearate.

A similar series of palmitic esters has been studied on aqueous surfaces by Adam (1929), and by Alexander & Schulman (1937). If the 'alcohol' chain contains four or fewer carbon atoms it is forced under the surface and vertically opposes the 'acid' chain in the condensed films. Whereas if the 'alcohol' chain contains more than four carbon atoms, the lateral adhesion between it and the 'acidic' chain is sufficient to bend the molecule up into the form of a hairpin. Alexander & Schulman obtained additional evidence in support of these orientations from measurements of the rate of hydrolysis, and of the vertical component of the dipole moment of the molecules in the films. In the case of cetyl palmitate they showed that the configuration in space of the polar group is so altered by the molecule being bent

into the form of a hairpin, that the vertical component of the dipole moment is reduced to zero.

The values of U suggest that the stearic esters are similarly oriented at the steel surface. In the case of methyl stearate, which has the smallest 'alcohol' radical attached to the —CO.O— group, the latter polar group would be able to approach most closely to the metal, which would account for the methyl ester having the greatest heat of adsorption. The larger ethyl and butyl radicals would hinder the close approach of the polar group to the surface, thus accounting for the low heats of adsorption of these esters. The rise in the heat of adsorption when the length of the 'alcohol' chain is increased to six carbon atoms suggests that, as in the case of the films on aqueous surfaces, the lateral adhesion between the 'alcohol' and 'acidic' chains is now bending the molecules into the form of a hairpin, bringing the polar group nearer to the surface again. If this were the case the heat of adsorption might not be as high as for the methyl ester since there might be some distortion of the polar group which would alter the component of its dipole moment resolved vertically to the surface. Also in taking up a position in which the least strain occurs, the orientation of the polar group with respect to the surface may be altered. Any reduction in the vertical component of the dipole moment due to these causes would be greater in the case of octadecyl stearate, in which the 'alcohol' and 'acidic' chains are almost equally long, which would account for the heat of adsorption of this substance being lower than that of the hexyl ester.

Glycol and glycerol stearates

Adam (1929) has shown that when films of the glycol esters on aqueous surfaces are compressed, the molecules become bent into the form of a hairpin with the hydrocarbon chains tightly packed together. The glycerol esters behave similarly on aqueous surfaces (Adam 1922). This must involve considerable strain in the molecule, since in the crystal lattice the glycerol esters exist as 'prong' shaped molecules, the middle hydrocarbon chain opposing the other two (Clarkson & Malkin 1934). If glycol and glycerol stearates are similarly oriented at a steel surface it might be expected that the heat of adsorption per —CO.O— group would be less for these esters than in the case of methyl stearate.

The results obtained for solutions of these esters are given in figure 6.

The values of the heats of adsorption are

glycol distearate, 10,500 cal./g.mol.; glycerol tristearate, 7,800 cal./g.mol.

The heat of adsorption per —CO.O— group is therefore 5300 cal. for glycol distearate and 2600 cal. for glycerol tristearate, compared with 8600 cal. in the case of methyl stearate. These results suggest that the glycerol and glycol esters are similarly oriented at aqueous and metal surfaces.

Octadecyl acetate

This ester differs from methyl stearate in having a long-chain 'alcohol' radical and a small 'acid' group. Since rotation is possible about single linkages it does not

necessarily follow that the resultant dipole moments of these two esters perpendicular to the surface will be very different due to this interchange of radicals. The results for solutions of octadecyl acetate are given in figure 6. The heat of adsorption is 7500 cal./g.mol. compared with 8600 cal./g.mol. in the case of methyl stearate.

1-nitro-octadecane

The results obtained for solutions of this substance are shown in figure 7. The heat of adsorption is:

13,000 cal./g mol.,

a high value which would be expected in view of the highly polar nitro group (Hunter & Partington 1933).

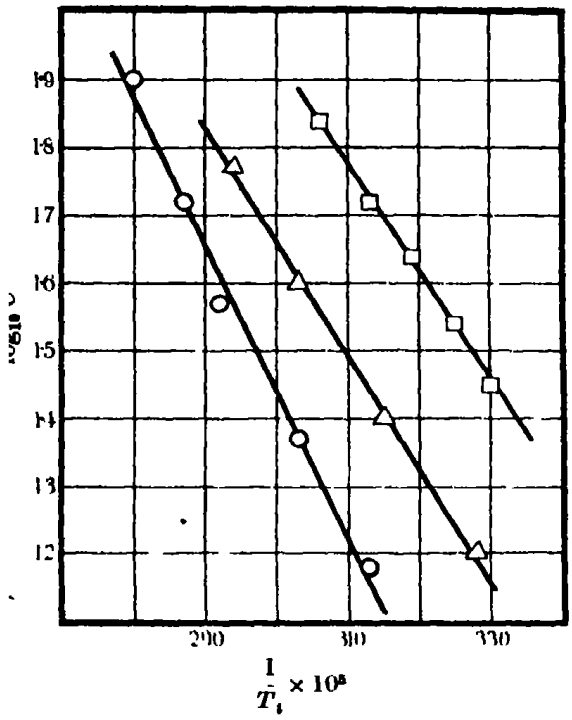


FIGURE 6. ⊙ glycol distearate; △ glycerol tristearate; [ ] octadecyl acetate.

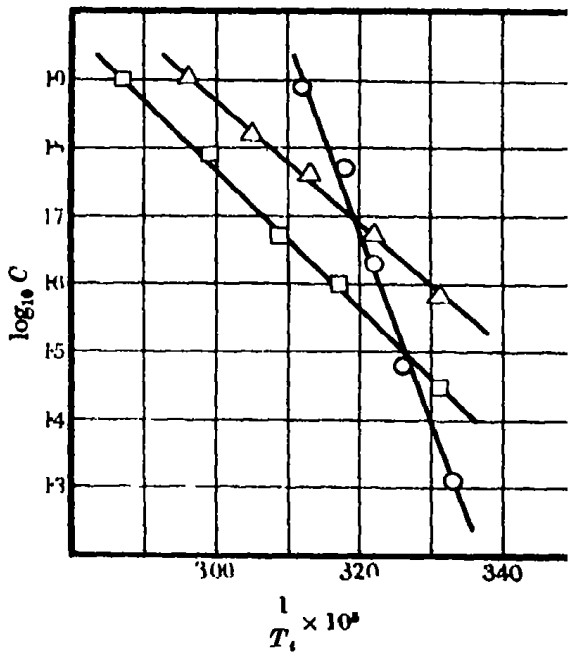


FIGURE 7. ⊙ 1-nitro-octadecane; △ octadecyl cyanide; □ octadecyl thiocyanate.

Octadecyl cyanide and thiocyanate

The results for solutions of these substances are also plotted in figure 7. The heats of adsorption are

octadecyl cyanide, 4300 cal./g.mol.; octadecyl thiocyanate, 4700 cal./g.mol.

These values are unexpectedly low in view of the high dipole moments: $3{\cdot}9 \times 10^{-18}$ e.s.u. for the —CN group and *ca.* $3{\cdot}6 \times 10^{-18}$ e.s.u. for the —SCN group (Hunter & Partington 1932), compared with *ca.* $3{\cdot}2 \times 10^{-18}$ for the $-NO_2$ group

(Hunter & Partington 1933). Parachor and surface measurements (cf. Alexander 1941) show that the —CN group occupies a much greater volume than would be expected from the atoms composing it. It is possible, therefore, that the negative charge on the nitrogen atom is unable to approach the metal surface closely enough for the attraction between this negative charge and the positive field of the metal to give rise to a strong adsorption.

Preparation of materials

Octadecyl alcohol

This material, from which all the derivatives of normal octadecane were prepared, was obtained by reduction of ethyl stearate (from stearic acid m.p. 69° C) as described by Adam & Dyer (1925). After recrystallization from ether it melted at 58° C.

Octadecyl chloride

Octadecyl alcohol was warmed with an excess of phosphorus pentachloride for 2 hr. on the water bath. The product was then poured into water and the octadecyl chloride extracted with 40/60 petroleum ether. The extract was washed with water and then shaken with concentrated sulphuric acid to convert any unchanged octadecyl alcohol into the sulphate, which was subsequently removed by shaking with a mixture of equal volumes of methyl alcohol and water. This mixture dissolves the sulphate, but not the chloride, and the presence of the methyl alcohol prevents the formation of troublesome emulsions. The petroleum ether solution was then dried over calcium chloride, the solvent removed and the residue distilled *in vacuo*. The fraction distilling between 180 and 190° C at 12 mm. was collected and redistilled. The product was a waxy substance melting sharply at 21° C.

Octadecyl bromide

This substance was prepared by passing dry hydrogen bromide into octadecyl alcohol at 110° C (cf *Organic syntheses*, **15**, 24). After fractionation *in vacuo* it melted at 28° C.

Octadecyl iodide

This material was prepared by the action of iodine on octadecyl alcohol in the presence of red phosphorus as described by Smith (1932) It was crystallized several times from a mixture of methyl alcohol and ethyl ether and melted at 33° C.

α-bromostearic acid

Stearic acid (m.p. 69° C) was treated with an excess of bromine in the presence of red phosphorus as described by Hell & Sadomsky (1891). After two crystallizations from 60/80 petroleum ether followed by four from benzene it melted sharply at 60° C. Hell & Sadomsky (1891) gave 60° C; and Le Sueur (1904) gave 60–61° C.

α-hydroxystearic acid

α-bromostearic acid was hydrolysed by refluxing for 24 hr. with a slight excess of dilute aqueous potash as recommended by Le Sueur (1904). After crystallization from ether followed by recrystallization from chloroform the product melted at 91° C. Le Sueur gave 91–92° C.

α-chlorostearic acid

Since direct chlorination would almost certainly result in the introduction of more than one chlorine atom in the α-position and probably others further up the hydrocarbon chain, it was decided to attempt the preparation of this substance from α-hydroxystearic acid. Several attempts were made by treating this substance with phosphorus pentachloride under various conditions: none was successful.

α-iodostearic acid

This material was readily obtained by refluxing α-bromostearic acid with an excess of potassium iodide in alcohol (Ponzio 1904, 1911) After recrystallization from benzene it melted at 67° C. Ponzio gave 66° C.

Stearic esters of n-alcohols, glycol and glycerol

The preparation, purification and properties of these esters have been described in a previous communication (Frewing 1942).

1-nitro octadecane

Freshly prepared and finely powdered dry silver nitrite was allowed to stand in contact with a solution of octadecyl iodide in 40/60 petroleum ether for 24 hr. After removal of the solvent the product was crystallized three times from methyl alcohol: it melted at 43·5° C.

Octadecyl cyanide

This substance was prepared by the action of alcoholic potassium cyanide on octadecyl iodide, as described by Levene & Taylor (1924). The product was recrystallized from methyl alcohol, and from a mixture of methyl alcohol and ethyl ether: it melted at 42° C. Levene & Taylor gave 42·5–43·5° C.

Octadecyl thiocyanate

The method used by Wheeler & Merriam (1901) in the preparation of cetyl thiocyanate was employed. This consisted in refluxing octadecyl iodide with alcoholic potassium thiocyanate and yielded a product with a melting point of 27° C, which was unchanged by recrystallization from methyl alcohol.

Discussion

Although the pure long-chain compounds which have been investigated contain polar groups of widely different structure and chemical activity, the frictional behaviour of the solutions of all these compounds in the neighbourhood of the transition temperature is essentially the same. The motion is at first smooth and then, as the transition temperature is reached, changes to stick-slips which increase in size with further rise of temperature. On cooling, this process is reversed. The temperature at which the transition occurs is independent of the number of times the surface is heated and cooled, provided that the temperature does not rise sufficiently to cause some chemical change—such as oxidation—in the oil or the solute. This general behaviour makes it very unlikely that the polar body is adsorbed by forming—as is sometimes supposed—a salt or other chemical compound with the steel surface.

The fact that the values of the coefficient of friction are the same as those previously obtained for monolayers of some of these long-chain compounds (Frewing 1942) suggests that the lubrication is under true boundary conditions and is effected by an adsorbed monolayer. It must be emphasized that the steel surfaces used in this investigation were prepared in air, and were therefore covered with the usual thin oxide and moisture films. Bowden & Hughes (1938, 1939) showed that when metal surfaces were thoroughly outgassed in a high vacuum, a unimolecular adsorbed layer of caproic acid was not sufficient to reduce the coefficient of friction to the low values usually associated with boundary lubrication, and these low values were not attained until a comparatively thick layer had condensed on the metal surfaces. The presence of the thin oxide film evidently reduces the field of force of the metal atoms in the surface, enabling boundary lubrication to be effected by unimolecular adsorbed films of long-chain compounds. It is possible that, with really clean surfaces, such as those studied by Bowden & Hughes, the fatty acids may be adsorbed by some chemical mechanism, since the forces of the metal atoms will not be partially saturated by an oxide film.

The variation of the transition temperature with concentration for all the compounds examined agrees well with the hypothesis that the adsorbed film is in equilibrium with the solution. When the plot of $\log_{10} C$ against $1/T_t$ did not give a straight line this discrepancy was traced to the presence of an impurity. For example, an old specimen of butyl stearate gave a curve which became steeper as the concentration decreased. This was due to the presence of a trace of free stearic acid resulting from slight hydrolysis of the ester. A freshly prepared specimen which was carefully purified to remove all trace of stearic acid gave the normal results recorded in figure 5.

Further support for the view that these long-chain compounds are adsorbed by the interaction of the dipoles in their polar groups with the metal atoms in the surface is afforded by the values of the heats of adsorption. Thus it is difficult to see how octadecyl chloride and 1-nitro octadecane could react chemically with the

surface, yet their heats of adsorption are as high as that of the fatty acids. And octadecyl chloride, which has a greater dipole moment, has also a greater heat of adsorption than octadecyl iodide. The values for the α-substituted stearic acids and the esters give further support for this mechanism of adsorption.

Although a fairly large concentration is necessary, in many cases, if the polar body is to build up a sufficiently complete film for smooth sliding to occur at temperatures such as would be obtained, even in the cooler parts of an engine, it does not necessarily follow that the addition of smaller amounts of these compounds, to the lubricating oil, would not be beneficial. The greater part of lubrication in practice occurs under conditions of 'fluid' or 'viscous' lubrication, that is to say, by comparatively thick films which are maintained between the moving parts by hydrodynamic forces. Beeck, Givens & Smith (1940) have shown that the addition to white oil of 1% of oleic acid enabled conditions of fluid lubrication to be attained, on their apparatus, at a lower critical speed, although the transition temperature of a solution of this concentration is below room temperature.

The author wishes to express his thanks to the Shell Refining and Marketing Company, Limited, for permission to publish this work, which was carried out in their laboratories.

References

Adam, N. K. 1922 *Proc Roy. Soc.* A, **101**, 516.
Adam, N. K. 1929 *Proc. Roy. Soc.* A, **126**, 366.
Adam, N. K. & Dyer, J. W. W. 1925 *J. Chem. Soc.* p. 70
Alexander, A E. 1941 *Trans Faraday Soc.* **37**, 1.
Alexander, A E. & Schulman, J. H. 1937 *Proc. Roy. Soc.* A, **161**, 115
Beeck, O , Givens, J W. & Smith, A. E. 1940 *Proc. Roy Soc.* A, **177**, 90.
Bowden, F. P. & Hughes, T P 1938 *Nature, Lond.*, **142**, 1039.
Bowden, F. P. & Hughes, T P. 1939 *Proc Roy. Soc.* A, **172**, 263
Bowden, F. P. & Leben, L. 1939 *Proc. Roy Soc.* A, **169**, 371.
Bowden, F. P. & Tabor, D. 1939 *Proc. Roy. Soc.* A, **169**, 391.
Bowden, F. P. & Tabor, D. 1942 *Bull. Comm. Australia C S I.R.* no. **145**.
Clarkson, C. E. & Malkin, T. 1934 *J. Chem Soc.* p. 666.
Frewing, J. J. 1942 *Proc. Roy. Soc.* A, **181**, 23.
Hell, C. & Sadomsky, J 1891 *Ber dtsch. chem Ges.* **24**, 2388.
Herman, R. C. 1940 *J. Chem. Phys.* **8**, 252.
Hughes, T. P. & Whittingham, G. 1942 *Trans. Faraday Soc.* **38**, 9.
Hunter, E. C. E. & Partington, J. R. 1932 *J. Chem. Soc.* p. 309
Hunter, E. C. E. & Partington, J. R. 1933 *J. Chem. Soc.* p. 2825.
Levene, P. A. & Taylor, F. A. 1924 *J. Biol. Chem.* **59**, 905.
Le Sueur, H. R. 1904 *J. Chem. Soc.* p. 827.
Malkin, T. 1931 *J. Chem. Soc.* p. 2796.
Ponzio, G. 1904 *Gazz. Chim. Ital.* **34**, II, 80.
Ponzio, G. 1911 *Gazz. Chim Ital.* **41**, I, 786.
Smith, J. C. 1932 *J. Chem. Soc.* p. 737.
Wheeler, H. L. & Merriam, H. F. 1901 *J. Amer. Chem. Soc.* **23**, 283.

The band spectrum of nitrogen: new singlet systems

By A. G. Gaydon, D.Sc.

Chemical Engineering Department, Imperial College, London, S.W. 7

(*Communicated by Sir Alfred Egerton, Sec.R.S.—Received* 20 *May* 1943)

[Plates 6, 7]

Five new systems due to N_2 have been observed in a mildly condensed discharge through nitrogen. Rotational analyses of some of the bands have been made, and it is shown that all correspond to transitions to the $a\,^1\Pi_u$ upper level of the Lyman-Birge-Hopfield System. An additional progression of the system studied by Van der Ziel, here referred to as the Fifth Positive System, has been found, necessitating a revision of the upper vibrational quantum numbers. Some of Kaplan's systems have also been examined, and the rotational analysis of one shows that it corresponds to a transition to the lower level of the Fifth Positive System. This level very probably lies a little below $a\,^1\Pi_u$ and must therefore be metastable; it may play an important part in the formation of active nitrogen. The rotational and vibrational constants for the singlet electronic states of N_2 are tabulated

Introductory and experimental

The investigation of the nitrogen spectrum described here was commenced with the object of obtaining further information about the heat of dissociation of N_2, the accepted value for which has recently been doubted by the author and Dr Penney (1942). Although this object has not, so far, been achieved, five new band systems of nitrogen have been discovered and analysed, and additional data and information has been obtained for other singlet systems of N_2 previously reported by Van der Ziel and by Kaplan.

The main singlet system of nitrogen, the Lyman-Birge-Hopfield System, lies down in the vacuum ultra-violet below 2000 A, while the two strong triplet systems, known as the First and Second Positive Systems, occupy the visible and near ultra-violet as far as about 2900 A. The region from 2000 to 2900 A is free from any really strong systems of N_2, and is therefore suitable for the study of weaker bands which lie in this region. With fairly vigorous conditions of excitation the single progression of bands discovered by Fowler and Rayleigh and known as the Fourth Positive System, a system of bands studied by Van der Ziel, which will later be referred to as the Fifth Positive System, and many weaker bands, degraded in either direction, can be observed. The chief difficulty in obtaining these bands is to eliminate the very strong γ system of NO which is very persistent, and was formerly known as the Third Positive System of N_2. With an uncondensed discharge at low pressure it seems almost impossible entirely to eliminate these bands; with a mildly condensed discharge, however, these NO bands are, as pointed out by Fowler and Rayleigh, less troublesome.

The discharge tube used for these experiments was about 30 cm. long, the central portion, about 20 cm. long, consisting of a capillary tube of 2 mm. bore. The dis-

charge was viewed end-on through a hollow cylindrical steel electrode and a thin fused quartz window. The discharge was maintained by an ordinary 10 in. induction coil with a small condenser and $\frac{1}{8}$ in. spark gap. A slow flow of chemically dry, oxygen-free nitrogen as supplied by the British Oxygen Co. was maintained through the tube at a pressure of a few mm. of mercury. All rubber tubing apart from one length of about 2 in. to the gas cylinder was eliminated, as it was found that introduction of rubber tended to bring up the NO bands. Under these conditions the spectrum was quite clean, only a trace of the strongest NO γ bands and the carbon line at 2478 A being observed as impurity. Most of the spectrograms were taken with a large Littrow-type quartz instrument (Hilger E. 1), but a few plates to cover the region between 2100 and 2240 A were taken with a medium-size (E. 2) quartz spectrograph. Details of the spectrograms reproduced in the plates are given at the end of the paper.

As may be seen from the plates, the region studied shows a very large number of bands of widely assorted types. In addition to the Fourth Positive bands, which are due to a $^3\Sigma \rightarrow {}^3\Pi$ transition, and the Fifth Positive bands, some of which have been studied in detail by Van der Ziel, many of these bands have been reported by Kaplan (1934, 1935), whose measurements are, however, only quoted to the nearest angstrom and are not sufficiently precise for certain identification or analysis.

Examination of the wave-number intervals between the heads of bands of similar type revealed a number of intervals of about 1666 cm.$^{-1}$. This agrees with the vibrational interval for the $a\,^1\Pi_u$ state, the upper level of the main Lyman-Birge-Hopfield bands. More detailed examination and rotational analysis of the bands showing this 1666 interval has revealed five new systems all having the $a\,^1\Pi_u$ state as their final level. The value of B_0 for the $a\,^1\Pi_u$ state from the analysis of the new bands comes out at between 1·615 and 1·60 This compares with the rather higher value of 1·632 obtained by Watson & Koontz (1934) from their analysis of bands of the main Lyman-Birge-Hopfield System. Watson & Koontz claim a fairly high accuracy for their value, but examination of their measurements shows that the value depends on the (0, 1) and (0, 4) bands, for the (0, 1) band the P and Q branches are superposed all the way; this is surprising, as the P branch would be expected to show a greater spacing than the Q branch and so should run through it; for the (0, 4) band the analysis is only fragmentary, and from such combinations of lines as can be made the values of B_0 so derived show a rather big scatter, ranging from 1·61 to 1·67. Watson & Koontz note that the $B_v - v$ curve obtained by combining their values of B_0 and B_1 with those for higher vibrational levels obtained by Appleyard (1932) shows a marked departure from linearity. It is probable that Watson & Koontz's value is less accurate than formerly supposed, and the most probable value for B_0 seems to be about 1·61 as obtained from the larger wave-number dispersion used in the present experiments

The detailed analysis of the five new systems is presented in the next five sections. The systems have been named the *P*, *Q*, *R*, *S* and *T* systems, and their upper

electronic states will be denoted by the letters p, q, r, s and t, these middle letters of the alphabet being selected to avoid confusion with other electronic states reported by Watson & Koontz. Later sections deal with the Fifth Positive System, with Kaplan's systems, and other bands, and finally the molecular constants for the known singlet states of N_2 are summarized and discussed.

System P, $p\,^1\Sigma_g^-$ to $a\,^1\Pi_u$

This system, although the last to be analysed and the least completely observed because of overlapping by the very strong Second Positive bands, is described first because its upper level is the lowest of the new states found Also it is probably the strongest of the new systems, although less obvious because of overlapping. The three bands of this system were first observed in a hollow cathode discharge through nitrogen and were provisionally assigned to N_2^+ on experimental grounds. Analysis of the (0, 0) band, the head of which is clear of overlapping structure, showed, however, that the band was due to a transition between singlet states and must therefore be ascribed to the unionized molecule

Three bands forming a single progression have been observed Their intervals agree with those for the $a\,^1\Pi_u$ state, and assuming that the initial vibrational quantum number is 0, the bands are therefore the (0, 0), (0, 1) and (0, 2). They are degraded to shorter wave-lengths and show a sharp P head and less marked Q head. The (0, 1) and (0, 2) bands are too overlapped by Second Positive structure to permit of detailed analysis, but the head of the (0 ,0) band is resolved into rotational fine structure and measurements are given in table 2. The analysis shows that the lines have the alternating intensities characteristic of homo-nuclear molecules, the members of the Q branch with odd J values and the R branch with even J values being the stronger. Since the band corresponds to a transition to a Π state and shows P, Q and R branches, the initial level must be Σ Since N_2 obeys Bose-Einstein statistics and it is the odd levels of the upper state which have the greater statistical weight, the initial state could be either $^1\Sigma_u^+$ or $^1\Sigma_g^-$. The fact that the transition is to $^1\Pi_u$ shows that it must be $^1\Sigma_g^-$.

For the (0, 0) band the lines of the P branch are crowded together into the strong first head and are unresolved. The R branch can be followed from R(4) to R(9), after which the structure is masked by the Second Positive band. The strongest branch is the Q branch which proceeds normally up to Q(10) after which there

Table 1. Vibrational scheme for system P

v', v''	0, 0		0, 1		0, 2
λ head	2827·1		2967·0		3118·6
ν head	35361·4	*1667*	33694	*1638*	32056
ν origin	35371·3				
B'	1·93				
B''	$1{\cdot}61_5$				

appears to be a perturbation, and assignments of lines to this branch beyond this point are uncertain. The vibrational scheme and constants are given in table 1, and the rotational analysis in table 2.

TABLE 2. INDIVIDUAL LINES IN THE NEIGHBOURHOOD OF THE (0, 0) BAND OF SYSTEM *P*. WAVE-NUMBERS OF THE LINES ARE GIVEN IN CM.$^{-1}$ FOLLOWED BY VISUAL ESTIMATE OF INTENSITY AND CLASSIFICATION TO P, Q, OR R BRANCH WITH *J* NUMBER WHERE POSSIBLE

ν	I	Classification	ν	I	Classification	ν	I	Classification
35361·4	8*H*	P 6	35388·9	6	Q 7	35415·4	4	?Q 11
372·0	3*H*	Q 1	393·8	4	Q 8	419·8	3	R 7
373·0	2	Q 2	397·0	3	R 4	422·2	3	?Q 12
375·1	4	Q 3	399·3	7	Q 9	428·6	6*b*	R 8
377·5	3	Q 4	403·9	4	R 5	431·6	2	
380·8	5	Q 5	405 8	3	Q 10	436·8	5*b*	
383·9	1		411 5	4	R 6	437·4	5*b*	R 9
384·4	4	Q 6				445·0	2	

H = head, *b* = broad, possibly double.

SYSTEM *Q*, $q^1\Pi_g$ TO $a^1\Pi_u$

Three bands in a single progression having heads at 2746 (0, 0), 2878 (0, 1) and 3020 A (0, 2) have been observed The last is strongly overlapped by other structure, and the (0, 1) has its 'tail' rather confused by lines of a Second Positive band, but a rotational analysis of this and the (0, 0) band has been made. Each band consists of two branches without alternating intensities, and the structure is consistent with a $^1\Pi \rightarrow {}^1\Pi$ transition with unresolved Λ doubling. For both bands the lines of the R branch, which close up to form a sharp head, are only partially resolved, but the P branches are well resolved and run smoothly up to P (12) after which there is some indication of a perturbation (presumably in the upper levels) and resolution into two Λ components, but overlapping by other structure and limited resolving power have prevented definite confirmation of the existence of this perturbation.

The vibrational scheme and molecular constants are set out in table 3, and the classification of lines of the (0, 0) and (0, 1) bands in table 4.

TABLE 3. VIBRATIONAL SCHEME FOR SYSTEM *Q*

v', v''	0, 0		0, 1		0, 2
λ head	2746 15		2877·93		3020·42
ν head	36403·8	*1666·8*	34737·0	*1638·7*	33098·3
ν origin	36394·5	*1666·7*	34727·8		
B'	$1{\cdot}36_5$		$1{\cdot}36_6$		
B''	$1{\cdot}60_5$		$1{\cdot}58_5$		

TABLE 4. INDIVIDUAL LINES OF THE (0, 0) AND (0, 1) BANDS OF SYSTEM Q. WAVE-NUMBERS IN CM.$^{-1}$. INTENSITIES ARE VISUAL ESTIMATES

(0, 0)					
ν	I	Classification	ν	I	Classification
36403·8	5H	R 5, R 6	36373 9	4	P 5
401·9	2?	R 8	368·1	4	P 6
400·3	2?	R 9	362·1	4	P 7
397·4	2	R 10	355·6	4	P 8
395·0	3	?R 11	348·3	4	P 9
393 8	1?		340·6	4	P 10
387·9	2	P 2	332·3	4d	P 11
383·6	2	P 3	323·2	3d	P 12
378·8	3	P 4	314·5	2d	?P 13
(0, 1)					
34737·0	5H		34675·7	3	P 10
727·3	3	R 12	667·8	3	P 11
724·1	3	R 13	659·9	3	P 12
720 8	3	P 2, R 14	659 3	2	
717·0	3	P 3, R 15	651 0	3	?P 13
712·5	3	P 4	644·4	2	
707·5	3	P 5	642·2	2	P 14
702·3	5	P 6	635·2	3d	
695·9	3	P 7	632·8	3d	P 15
689·7	3	P 8	622·7	2d	P 16
683·0	3	P 9	612 2	3d	P 17

H = head consisting of several unresolved lines, d = diffuse line.

SYSTEM R, $r\,^1\Sigma_g^-$ TO $a\,^1\Pi_u$

The (0, 0) band of this system at 2671 A is the most outstanding of the new bands. The band is very slightly degraded to shorter wave-lengths, the strong Q branch being piled up into a single very broad line with edges at 2671·2 and 2671·7 A. The P and R branches are well resolved and show clearly alternating intensities. Details of the rotational analysis of this band are given in table 6. The lower state is again found to be $a\,^1\Pi_u$, and since it is the members of the P and R branches with even J (that is odd J') which are stronger it follows that the initial state is $^1\Sigma_g^-$.

It will be observed that both the P and R branches show an abrupt decrease in intensity at high J value. This might be due to underlying weaker band structure but more probably indicates predissociation in the upper state, an explanation which would also account for the failure to observe any bands from higher vibrational levels

Other bands of this system which might be present are masked by the stronger Second Positive bands, but the piled up Q branch of the (0, 1) band at 2796 A is clearly visible among the line structure of the tail of the (4, 1) Second Positive band and serves to give additional proof that the final state is $a\,^1\Pi_u$. The heads and molecular constants are given in table 5.

TABLE 5. VIBRATIONAL SCHEME FOR SYSTEM *R*

v', v''	0, 0		0, 1
λ Q head	2671·67		2796·0
ν Q head	37417·7	*1664*	35754
ν origin	37416·9		
B'	1·67		
B''	1·60		

TABLE 6. INDIVIDUAL LINES OF THE (0, 0) BAND OF SYSTEM *R*

ν	I	Classification	ν	I	Classification	ν	I	Classification
37470 7	4		37439·2	3	R 5	37402·5	3	P 5
469·8	4		435 5	4	R 4	400·2	5	P 6
467·0	4	R 12	431·1	2	R 3	398·0	3	P 7
463·9	2	R 11	427 3	2	R 2	395·7	5	P 8
460·8	4	R 10	425 2		Q branch	393·6	2	P 9
456·5	2	R 9	417 7			391·5	4	P 10
452·6	4	R 8	410 1	4	P 2	390 0	2	
447·7	2	R 7	407 6	2	P 3	389·0	2	?P 11
443·7	5	R 6	404 9	5	P 4	387·1	3	?P 12

SYSTEM *S*, $s\,{}^1\Sigma_g^-$ TO $a\,{}^1\Pi_u$

The bands of this system are very similar to those of the *R* system, but are slightly degraded to the red. The (0, 0) band is most clearly developed and shows a weak R head at 2395·12 A and a sharp Q head at 2397·08. The P and R branches show clear alternation of intensities It may be noted that for this, and also for the *T* system, the lines of the P branch are considerably stronger than the lines of the R branch, it is to be expected that for a Σ to Π transition the R lines should be slightly weaker than the P lines, but for these systems the difference appears qualitatively rather larger than would be expected; a similar effect has been observed for NiH (Gaydon & Pearse 1935).

The rotational analysis of the (0, 0) band, for which data are presented in table 8, shows that the transition is again from a $^1\Sigma_g^-$ state to $a\,{}^1\Pi_u$, the value for B_0'' coming out at about 1·605. The (0, 1) band shows a strong Q head, but the band is overlapped by other structure and has not been analysed. The (0, 2) band shows a line-like Q head at 2603·3 A with a P branch showing good alternation of intensities. The R

TABLE 7. VIBRATIONAL SCHEME FOR SYSTEM *S*

v', v''	0, 0		0, 1		0, 2
λ Q head	2397·08		2496·78		2603·34
ν Q head	41704 6	*1665·2*	40039·4	*1638·6*	38400·8
ν origin	41705·4				38399·9
B'	$1·58_5$				$1·58_5$
B''	$1·60_5$				$1·56_5$

branch is so weak that it is probably invisible on the enlargement and although just visible on the negative is too weak to measure. The head of this band falls very near that of the strongest Vegard-Kaplan band (0, 5) at 2603·8, but the two bands are quite definitely not the same, the Vegard-Kaplan bands do not appear under the conditions of these experiments.

For the (0, 0) band the usual P, Q, R relationships check up fairly well, indicating that the Λ doubling in the $a\,^1\Pi_u$ state is small, although there is evidence that it is perhaps not quite negligible at higher J values, the Q lines probably going to the Λ component with the lower energy. For the upper state the rotational energy obeys the usual formula $B.J(J+1)$ quite well with a value for B of about 1·585 up to $J = 8$, but at higher J values the apparent value of B falls off rapidly. This may be seen readily in the P branch of the (0, 2) band for which the spacing increases rapidly at higher J values. It is probably better to regard this as due to a perturbation than to endeavour to represent it by an abnormally large constant D.

TABLE 8. INDIVIDUAL LINES OF THE (0, 0) AND (0, 2) BANDS OF SYSTEM S

0, 0 band						0, 2 band		
ν	I	Classification	ν	I	Classification	ν	I	Classification
41738·8	3	R head	41692·2	4	P 4	38400 8	10	Q branch
737·4	3		689·4	4	P 5, Q 15	400 0	10	
737·1	3		685·5	5	P 6	393·0	3	P 2
734·2	1?	R 9	682·1	3	P 7	387·2	3	P 4
732·2	3	R 8	678·3	5	P 8	384·7	1	P 5
729·7	2	R 7	675·1	3	P 9	381·3	3	P 6
727·1	3	R 6	671·1	5	P 10	379 1	1	P 7
724·3	1	R 5	667·2	3	P 11	376·0	3	P 8
720·6	2	R 4	662·6	5	P 12	372 8	1	P 9
718·2	1	R 3	657·8	4	P 13	369·7	3	P 10
714·5	0?	R 2	652·1	5	P 14	366·2	1	P 11
704·6	10	Q head	645 6	4	P 15	362·8	2	P 12
703·3	3?	Q 9	638·2	4	P 16	358 8	1	P 13
701·4	3	Q 11	631 0	3		354·3	2	P 14
699·1	3	Q 12	629 4	3		348 9	1	P 15
697·2	3	P 2, Q 13	619·2	3		342·6	2	P 16
693·6	1?	Q 14						

SYSTEM T, $t\,^1\Sigma_g^-$ TO $a\,^1\Pi_u$

This system closely resembles the two previous systems, and the rotational analysis of two of the bands indicates that the transition is again from a $^1\Sigma_g^-$ state to $a\,^1\Pi_u$. The band at 2281 A, which will provisionally be referred to as the (0, 0), has not been analysed although its structure is quite clear. Details of the (0, 1) and (0, 3) bands are given in table 10. The (0, 2) band, if present at all, is weak and masked by a Fifth Positive band. The Q head of the (0, 4) band is faintly discernible among the structure of a fairly strong Fifth Positive band.

The vibrational scheme and molecular constants are set out in table 9. The intensity distribution of the single progression, showing strong (0, 0), (0, 1) and (0, 3) but no (0, 2) band is unusual. There can be no doubt about the assignment of the values of v'', but the value of v' is less certain. If, as assumed here, $v' = 0$, then it must be presumed that the wave-functions for $v' = 0$ and $v'' = 2$ are so placed relatively that they integrate to practically zero.

This system, as given, has its initial state 2111 cm.$^{-1}$ above that of the S system. This at first sight appears to be a possible value for a vibrational separation The assumption that the T system was the $(1, v'')$ progression of the S system would have the additional advantage that the combined system would have a normal intensity distribution. However, the value of 2111 seems rather large for ω_i for a state with B around 1·6, and moreover the T system has a B' of 1·63 against 1·585 for S, this giving a large negative value for the constant α. It does not seem possible with existing knowledge to say whether S and T are really one or two systems. It seems perhaps best to treat the two progressions separately

TABLE 9 VIBRATIONAL SCHEME FOR SYSTEM T THE Q HEADS OF THE BANDS ARE GIVEN, THE LETTERS R AND V INDICATING THAT THE EDGE REFERRED TO IS DEGRADED TO LONGER OR SHORTER WAVE-LENGTHS

v', v''	0, 0		0, 1	0, 2	0, 3		0, 4
λ head			R 2371·59		R 2569 46		
	R 2281 48		V 2371·72		V 2569 79		V 2678·7
ν head	R43817 6	*1667 0*	V42150·6	—	V38902 1	*1581 7*	V37320·4
ν origin			42150·3		38901·4		
B'			1 63		1 63		
B''			1 585		1 55		

TABLE 10. INDIVIDUAL LINES OF THE (0, 1) AND (0, 3) BANDS OF SYSTEM T

(0, 1) band						(0, 3) band		
ν	I	Classification	ν	I	Classification	ν	I	Classification
42181·7	1	R 8	42141·4	2	P 3	38907·0	6	Q branch
178 5	2	R 7	138·4	2	P 4	902 1	6	
175·2	1	R 6	135 9	1	P 5	898·4	0	P 1
171·6	0	R 5	132·9	3	P 6	895 3	2	P 2
168·1	2	R 4	130·4	1	P 7	893·0	0	P 3
164·6	1	R 3	127·6	3	P 8	890 1	2	P 4
160·4	0	R 2	124 6	1	P 9	888·0	0?	P 5
152·9	10	Q branch	121·5	3	P 10	885·5	2	P 6
150·6	10		117 5	1*d*	P 11	880·7	2	P 8
147·0	1	P 1	112·4	2*d*	P 12	878 9	2*d*	P 9
144·4	2	P 2	105·9	1*d*	P 13	876·1	2	P 10
			097·3	1*d*	P 14	872·9	0	P 11
						868·9	1	P 12

d = diffuse.

As with the other two systems of similar structure, the rotational levels of the upper state behave normally at low J values, but at high J values the apparent value of B' falls off rapidly. This is very clearly seen in the plates on examination of the P branches of the (0, 0) and (0, 1) bands in which the spacing increases sharply at high J values There is also some signs that lines of high J value are diffuse, perhaps indicating predissociation.

The Fifth Positive or Van der Ziel's System

The first record of these bands of which the author is aware is that by Duncan (1925), who recorded several apparently new band systems of nitrogen in a controlled electron source. Duncan's experiments were, however, marred by the inclusion of bands due to impurities (CO and CO^+), and his paper has been largely neglected. Nevertheless, one of his systems is undoubtedly identical with this Fifth Positive System and includes, in addition to nearly all the bands listed by Van der Ziel, several bands of the new progression reported here. Appleyard (1932) independently reported two bands degraded to the violet at 2112·1 and 2033·6 A showing alternating intensities and therefore presumably due to N_2. The first detailed study of the bands was made by Van der Ziel (1934) by whose name the system has since frequently been referred. In view of the lack of priority in the discovery and the modification to the analysis proposed here it seems perhaps fairer to refer to the bands impartially as the Fifth Positive System of nitrogen, a name which aptly describes their condition of occurrence, the bands coming up under very similar excitation conditions to the Fourth Positive System, but less strongly.

Van der Ziel arranged the bands into two progressions and made a rotational analysis of three bands from which he found that the transition was between two Σ states, the odd J levels of the upper state and the even J values of the lower state having the greater statistical weight, so that the transition was either ${}^{1,3}\Sigma_u^+ \rightarrow {}^{1,3}\Sigma_g^+$ or ${}^{1,3}\Sigma_g^- \rightarrow {}^{1,3}\Sigma_u^-$. The present work on one of Kaplan's systems (see next section) shows that the transition is between singlet, not triplet states, and the transition must therefore be either ${}^1\Sigma_u^+ \rightarrow {}^1\Sigma_g^+$ or ${}^1\Sigma_g^- \rightarrow {}^1\Sigma_u^-$.

The vibrational intensity distribution of the two v' progressions observed by Van der Ziel is far from normal and the bands cannot be made to conform to the usual type of Franck-Condon parabola, a fact which was very difficult to explain and which had been commented on by Kaplan. Examination of the new bands observed in the present investigations revealed another progression of bands showing the same intervals as the lower state of Van der Ziel's bands. This progression obviously forms part of the system and is indeed the $v' = 0$ progression, those recorded by Van der Ziel being the $v' = 1$ and $v' = 2$ progressions. The complete system now has a normal intensity distribution and the vibrational scheme is set out fully in table 11, the wave-lengths with visual estimates of intensity in parentheses, wave-numbers to the nearest cm.$^{-1}$ and wave-number intervals (in italics) being given. In addition to the new progression the table includes measurements of

two new bands of the old progressions. It should be noted that, for convenience of presentation, the v' and v'' axes are not set the conventional way round.

In order to confirm the analysis of the new bands and to find the rotational constants it seemed desirable to make a rotational analysis of at least one band of

TABLE 11. THE FIFTH POSITIVE SYSTEM OF N_2

v'' \ v'	0		1		2
0	2198·9 (4)		2112 1 (5)		2033·6 (–)
	45463	*1867*	*Z*47330	*1828*	*Z*49158
	1506		*1504*		*1509*
1	2274·2 (6)		2181·5 (4)		2098 (2)
	43957	*1869*	45826	*1823*	47649
	1483				*1477*
2	2353·6 (4)		*m*		2165·1 (5)
	42474				46172
					1459
3	*m*		2331 0 (2)		2235·8 (3)
			42886	*1827*	44713
			1435		
4	2525·6 (2)		2411·7 (7)		
	39582	*1869*	*Z*41451		
	1415		*1413*		
5	2619·3 (4)		2496 9 (3)		
	38167	*1871*	*Z*40038		
			1388		
6			2586 5 (7)		2469 7 (4)
			*Z*38650	*1828*	*Z*40478
			1364		*1367*
7			2681 2 (5)		2556 0 (1)
			*Z*37286	*1825*	*Z*39111
			1346		*1344*
8			2781 6 (3)		2647 0 (2)
			*Z*35940	*1827*	*Z*37767
					1321
					2743·0 (1)
					*Z*36446

Z preceding the wave-number indicates measurement from Van der Ziel, not remeasured by author.

m indicates band masked by Fourth Positive band.

* indicates present, but head masked by other structure.

the new $v' = 0$ progression. Unfortunately, most of these bands are rather confused by overlapping structure from other bands. However, the (0, 1) band has been measured and although the measurements include some additional structure due to other bands the analysis proves quite convincingly that the band is of the expected type, the values of B' and B'' coming out at 1·735 and 1·45, and it is clear

that the R lines with even J numbering are the stronger, in agreement with expectation. The measurements and assignments are given in table 12.

TABLE 12. WAVE-NUMBERS AND INTENSITIES OF STRUCTURE BETWEEN 43955 AND 44040 CM.$^{-1}$ WITH ASSIGNMENT OF LINES TO THE (0, 1) FIFTH POSITIVE BAND. UNRESOLVED STRUCTURE IS BRACKETED TOGETHER

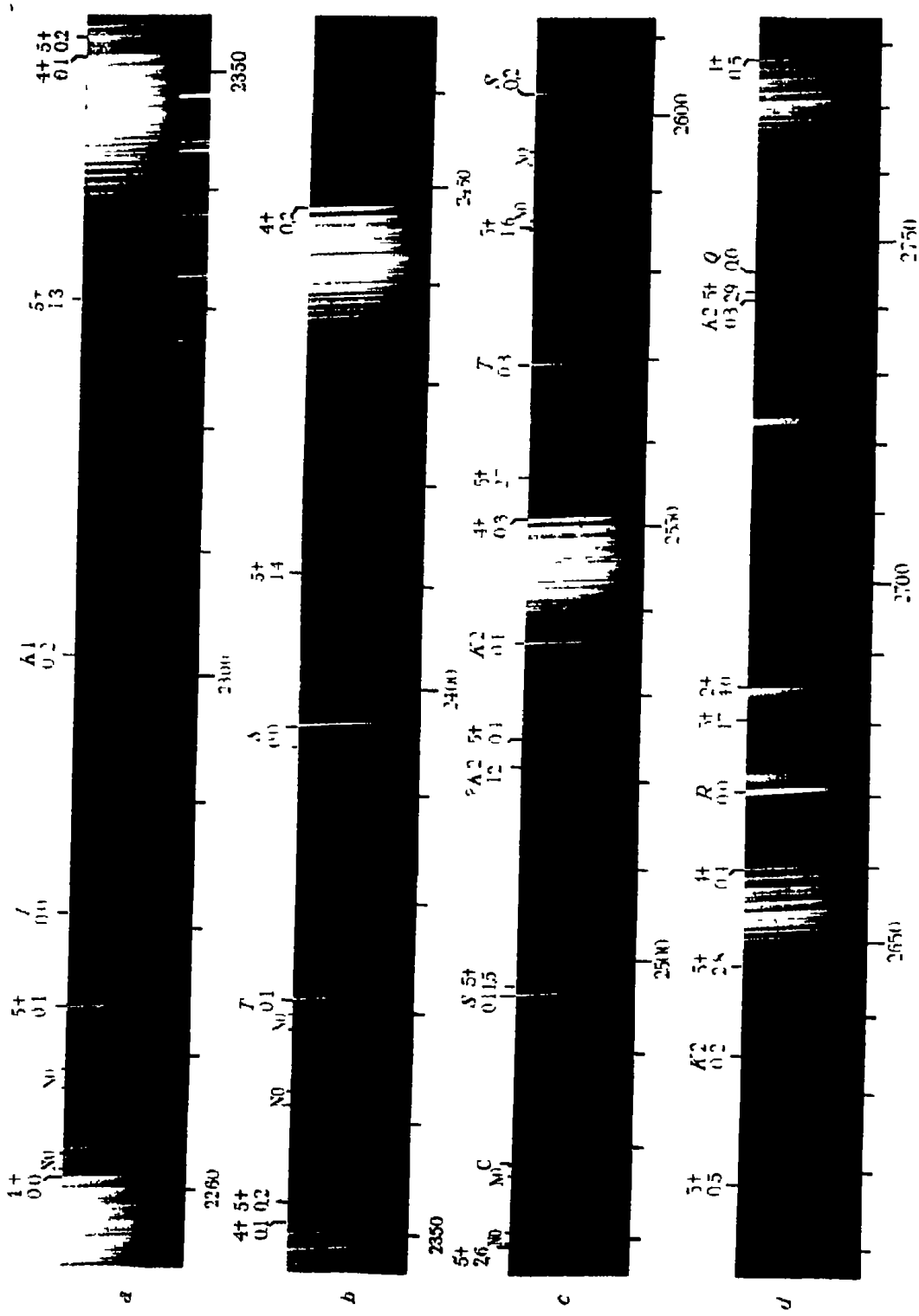
ν	I	Classification	ν	I	Classification	ν	I	Classification
43956·9	6	P5, P6	43973·1	1	?R1	44001·7	3*b*	R6
957·7	6	P7, P8, P4	977·7	4	R2	006·1	1	
960·1	3	P9, P2	982 9	2	R3	010·0	3*b*	R7
962 1	3	P10	986·9	1		015·5	3*b*	
964·3	1	?P11	988·8	4	R4	017·3	3*b*	R8
967·9	3	P12	992·5	1*b*		025·8	2	R9
970·8	1		995·2	2*b*	R5	035·2	3	R10

b = line broad or diffuse.

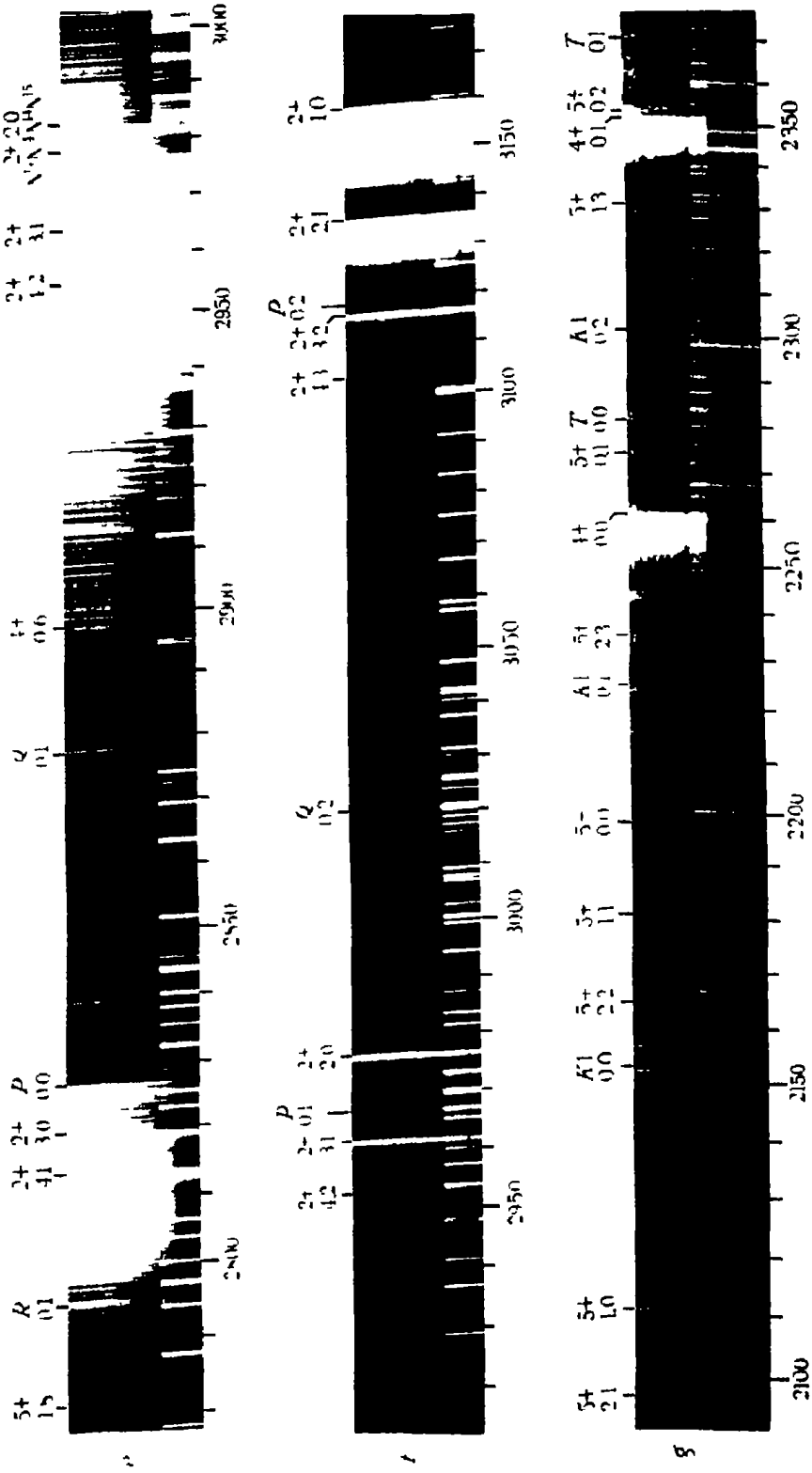
KAPLAN'S SYSTEMS

Kaplan (1934, 1935) has reported three rather weak band systems of N_2 in emission. He has published wave-lengths (to the nearest angstrom) and provisional vibrational analyses for these systems which indicates that their final electronic level may be identical with that of what we are now calling the Fifth Positive System. He also tentatively identified the wave-number interval between the upper levels of his bands and Van der Ziel's bands with intervals in Hopfield and Birge's Rydberg series of N_2 Van der Ziel (1937) expressed doubts about the analysis and identification. Some, although not all, of these bands observed by Kaplan have now been obtained by the author, and the much greater dispersion used has enabled rather more definite conclusions to be reached. While the identification with the Rydberg series interval is ruled out by the discovery of the new Fifth Positive progression, and can also be dismissed for the reasons given by Van der Ziel, it has been shown that certainly one system, and very probably two, do correspond to transitions to the lower state of the Fifth Positive bands.

Kaplan's First System. The (0, 1) and (0, 2) bands of this system are fairly well developed and the latter has been photographed on the E. 1 and a rotational analysis has been made There is a weaker head at 2153·6 A which fits quite well into the scheme, but this band is largely masked by a much stronger band of unknown origin lying to slightly shorter wave-lengths. The (0, 3) band is definitely present and the alternating intensities of the lines of the Q branch are visible, but the head is masked by other structure and cannot be measured. The bands of this $v' = 0$ progression are set out in table 13. Kaplan also records the bands (1, 3) at 2288 and (1, 4) at 2366 A. It is difficult to confirm or deny the presence of these bands on the author's plates. There is certainly a sort of head at 2366·9 A but it is partly obscured by NO


a
b
c
d
2350
2260
2300
2400
2350
2500
2450
2600
2550
2700
2650
2750

(*Facing p* 206)



and it is difficult to say if it is of the right type. It would lead to a value of about 1705 cm.$^{-1}$ for ω_1' which seems of the right magnitude.

TABLE 13. KAPLAN'S FIRST SYSTEM

v', v''	0, 0		0, 1		0, 2
λ head	2153·6		2225·8		2301·9
ν head	46420·0	*1506 0*	44914·0	*1485 6*	43428·4
ν origin					43435·6
B'					1·80
B''					1·44

The rotational analysis of the (0, 2) band, for which details are given in table 14, reveals that the band has P, Q and R branches with alternating intensities, the lines with even J having the greater strength The vibrational separations, the value for B_1'' and the fact that it is the even J levels which have the greater statistical weight all prove that the final state is the same as that involved in the Fifth Positive bands, so that the transition must be either $^1\Pi_u \rightarrow {}^1\Sigma_g^+$ or $^1\Pi_g \rightarrow {}^1\Sigma_u^-$

TABLE 14. INDIVIDUAL LINES OF THE (0, 2) BAND OF KAPLAN'S FIRST SYSTEM

ν	I	Classification	ν	I	Classification	ν	I	Classification
43428·4	3*H*	P 4, P 5	43445 0	1		43476 2	2	R 6
429 4	3	P 3, P 6	446·5	2	Q 5	482 4	2	Q 11
431·1	3	P 2, P 7	448 9	2	R 2	484·0	1	R 7
433·5	1	?P 8	450 7	3	Q 6	489·1	1	
436·7	2*H*	Q 1	454·7	2	R 3	491 4	2	Q 12
437·7	1	Q 2	455·3	2	Q 7	500 6	1*b*	Q 13
439·0	1	R 0	460 8	5*b*	Q 8, R 4	510 9	2	Q 14
439·9	2	Q 3	467·1	2*b*	Q 9	522·0	1*b*	Q 15
442·9	2	Q 4	474·5	5	Q 10	533·2	1	Q 16

b = broad or diffuse line; *H* = head with lines in neighbourhood barely resolved.

Kaplan's Second System The (0, 1), (0, 2) and (0, 3) bands of this system have been observed. The (0, 0) band is obscured by a Fourth Positive band. The (1, 3) band given as 2620 A falls exactly on the new (0, 5) Fifth Positive band, which it may possibly serve to strengthen. The only other band of the several reported by Kaplan is a head at 2522·3 A which *may* be the (1, 2) band. The vibrational scheme for the four bands which have been observed is set out in table 15. Although the (0, 2) band is free from overlapping bands, the rotational analysis has proved too difficult with the dispersion used, and although some sort of fit to a $^1\Pi \rightarrow {}^1\Sigma$ transition has been obtained it seems risky to comment on the nature of this band system, except to point out that the vibrational intervals agree as well as may be expected with those for the lower state of the Fifth Positive bands.

TABLE 15. BANDS OF KAPLAN'S SECOND SYSTEM

v' \ v''	1		2		3
0	2536·6		2636 2		2742·0
	39410		37923		36459
		1487	*1711*	*1464*	
1			2522·3		
			39634		

Kaplan's other systems. Of Kaplan's Third System consisting of bands at λλ 2472, 2392, 2316 and 2242 there is no very obvious sign. Kaplan also reported four progressions λλ 2428, 2360, 2297, 2236, λλ 2459, 2367, 2280, 2198, λλ 2272, 2204, 2139, and λλ 2477, 2391, 2309. Some of these bands coincide within the accuracy of his measurements with bands now definitely assigned to the Fifth Positive System or the new systems, and other bands may be present, but the present author is unable to confirm the reality of any one of these progressions. The conditions of excitation used by the author differ somewhat, of course, from those under which Kaplan observed his various systems.

OTHER BANDS

As can be seen from the plates, there is still a lot of band structure which is unaccounted for. It is not proposed to discuss this structure in detail as this would take up a lot of space and would give little more information than that obtainable from a careful study of the plates Some of the bands show fairly definite alternation of intensities, and the comparative freedom from impurities suggests that most if not all of this additional structure is to be attributed to N_2.

Among the more outstanding features are bands degraded to the red at 2424, 2498, 2524 and 2828 A, these being different in character but all showing fairly open rotational structure suggesting that singlet states are involved. Some of the bands degraded to the red in the shorter wave-length region may be an extension of Watson & Koontz's (1934) N_2^+ system. It also seems possible that a weak second progression to the Fourth Positive System might account for some of the heads degraded to shorter wave-lengths.

Attempts to analyse various parts of this unaccounted for structure have not given any convincing analyses. Long exposures with a rather more intense source on greater dispersion seem called for, and it is unlikely that further work on the singlet systems will be undertaken by the author. The work should not, however, prove very difficult, given time and adequate dispersion, and might well be of considerable value as it might settle the relative positions of the levels involved in the Fifth Positive bands and the fixed singlet states and might (see later) thus assist in the interpretation of the nitrogen afterglow.

Two new systems of higher multiplicity (probably triplet) due to N_2 have recently been reported by the author (1943) using discharges at high gas pressure. Improved plates of these systems have now been taken and it is hoped to describe them more fully in due course.

Discussion and conclusions

The analyses of the five new systems have given details about the nature and energy of five new electronic states of N_2. Four of the new states are of the type $^1\Sigma_g^-$, and it was at first thought these states might form an approximate Rydberg series. Attempts to fit them to such a series have, however, not been successful. The upper three states can readily be forced to fit approximately to a series, but as the series limit lies well below the ionization potential of N_2 it cannot be real.

Hitherto little has been known about the amount of energy required to excite the Fifth Positive bands. It may be noted that Duncan (1925) found bands, now shown to be identical with the Fifth Positive System, appearing in his controlled electron source at 15·5 V (125,000 cm.$^{-1}$); this probably represents an upper limit rather than the actual excitation potential From the present experiments and a study of the conditions used by Van der Ziel and by Kaplan, it seems that the Fifth Positive bands appear rather more easily, that is, with lower excitation, than the new systems. This seems to indicate fairly definitely that the upper level of the Fifth Positive System is not so very high, probably lying rather below 105,000 cm.$^{-1}$. This brings the lower state definitely below $a\,^1\Pi$. If this is so, as seems fairly certain, then this state must be metastable.

Attempts to explain the nitrogen afterglow have usually taken account of the metastability of the $A\,^3\Sigma_u^+$ state of N_2 and of the ^2D and ^2P states of the nitrogen atom There are several difficulties to be met in these explanations Absorption of the First Positive bands by active nitrogen has never been observed. The fact that the Vegard-Kaplan bands ($A\,^3\Sigma \rightarrow X\,^1\Sigma$) appear in emission at atmospheric pressure in an ozonizer discharge does not seem in agreement with the $A\,^3\Sigma$ state having the very long life required for its participation in the afterglow. The afterglow itself does not emit the Vegard-Kaplan bands. The strength of the afterglow bears little relation to the strength of the First Positive bands, ending on $A\,^3\Sigma$, in the exciting discharge. Also Herbert, Herzberg & Mills (1937) have failed to detect the presence of ^2D or ^2P nitrogen atoms by absorption through active nitrogen If, as now seems fairly certain, the $v\,^1\Sigma$ lower state of the Fifth Positive bands is metastable, then this might play an important rôle in active nitrogen. At least it can be said that the conditions for the production of the afterglow and the Fifth Positive System, namely, a mildly condensed discharge, are very similar.

As already stated, the Fifth Positive bands correspond either to the transition $^1\Sigma_u^+ \rightarrow {}^1\Sigma_g^+$ or to $^1\Sigma_g^- \rightarrow {}^1\Sigma_u^-$. It is possible to put forward arguments in favour of either alternative, but at present a clear decision does not seem possible. In either case some transitions linking the levels involved in the Fifth Positive or Kaplan's

Systems with levels of known energy should occur in accessible regions of the spectrum. Some of the unassigned band structure visible on the plates may be of this type, and it is unfortunate that analyses of these bands do not at present appear possible.

In table 16 the molecular constants for the singlet levels of N_2 are collected. The lower level of the Fifth Positive is denoted as v, and the upper levels of this system and Kaplan's First and Second Systems as y, w and x respectively.

TABLE 16. SUMMARY OF CONSTANTS FOR SINGLET STATES OF N_2

	State	ν_0	B_e	B_0	α	ω_e	ω_1	$x_e\omega_e$
t	$^1\Sigma_g^-$	112774·2		1·63				
s	$^1\Sigma_g^-$	110662·0		$1\ 58_5$				
r	$^1\Sigma_g^-$	106373 5		1·67				
q	$^1\Pi_g$	105351·1		$1·36_5$				
p	$^1\Sigma_g^-$	104327·9		1·93				
c	?$_u$	104394 6 104318·7		1·159				
b'	$^1\Sigma_u^+$	103677·7	1 145	1·144	0·002	$\omega_0 = 758·6$		4·16
b	?$_u$	101469 2						
y	$^1\Pi_u$ or $^1\Pi_g$	$v+46420$		1 80			(1705)	
x	$^1\Sigma_u^+$ or $^1\Sigma_g^-$	$v+45463$	1·74	$1·73_5$	$0\ 01_5$	1910	1869	20·5
w	?	$v+40914$					(1711)	
a	$^1\Pi_u$	68956 6	1 62	1·61	0·018	1694·7	1666·7	14 0
v	$^1\Sigma_g^+$ or $^1\Sigma_u^-$	?~60000	1 48	1·47	$0·01_5$	1527	1504	11 5
X	$^1\Sigma_g^+$	0		1·998	(0 018)	2359·6	2330 7	14 44

In the table the more directly observable quantities B_0 (the rotational constant for the lowest vibrational level) and ω_1 (the energy difference between the $v = 0$ and the $v = 1$ levels) are probably more accurate than the extrapolated values B_e and ω_e for zero vibrational energy. Data for the levels c and X have been taken from Sponer (1935). The b' state has been examined by Chulanovskii (1935). In addition to the levels given in the table, Watson & Koontz (1934) have reported normal state progressions in emission which suggest levels known as g, f and h at about 108937, 110192 and 112760 cm.$^{-1}$, their progressions d and e are believed by Chulanovskii to be part of the $b' \rightarrow X$ system. Hopfield (1930) and Worley & Jenkins (1938) have also reported Rydberg series.

In conclusion I wish to express my thanks to Sir Alfred Egerton for his keen interest in the work, to Dr R. W. B. Pearse for much helpful discussion and to the Council of the Royal Society for financial assistance.

References

Appleyard, E. T. S. 1932 *Phys. Rev.* **41**, 254.
Chulanovskii, V. M. 1935 *C.R. Acad. Sci. U.R.S.S.* **3**, 155
Duncan, D. C. 1925 *Astrophys. J.* **62**, 145
Gaydon, A. G. 1943 *Nature, Lond.*, **151**, 167.
Gaydon, A. G. & Pearse, R. W. B. 1935 *Proc. Roy. Soc.* A, **148**, 321.
Gaydon, A. G. & Penney, W. G. 1942 *Nature, Lond.*, **150**, 406.
Herbert, W. S., Herzberg, G. & Mills, G. A. 1937 *Canad. J. Res.* A, **15**, 35
Hopfield, J. J. 1930 *Phys. Rev.* **36**, 789
Kaplan, J. 1934 *Phys. Rev.* **46**, 534 and 631.
Kaplan, J. 1935 *Phys. Rev.* **47**, 259.
Sponer, H. 1935 *Molekulspektren.* Berlin: J. Springer.
Van der Ziel, A. 1934 *Physica*, **1**, 513
Van der Ziel, A. 1937 *Physica*, **4**, 373
Watson, W. W. & Koontz, P. G. 1934 *Phys. Rev.* **46**, 32
Worley, R. E. & Jenkins, F. A. 1938 *Phys. Rev.* **54**, 305

Description of Plates 6 and 7

(*a*) Spectrum of mildly condensed discharge through N_2. Large quartz (E.1) spectrograph. Ilford Q II plate. Exposure 11½ hr.

(*b*), (*c*) and (*d*) Mildly condensed discharge (E.1). Zenith plate. 1½ hr.

(*e*) Mildly condensed discharge (E.1). Zenith plate. ½ hr.

(*f*) Hollow cathode discharge in N_2 at low pressure (E.1). Zenith plate. ½ hr.

(*g*) Mildly condensed discharge. Medium quartz spectrograph (E.2). Ilford Q II plate. Exposure 6 hr.

An iron arc comparison spectrum is shown below each N_2 spectrum, and an approximate wave-length scale is marked below.

All bands definitely assigned to N_2 are marked above with an abbreviation for the system and the vibrational quantum numbers (v', v''). The following are the abbreviations

P, *Q*, *R*, *S*, and *T* for the new systems.

2+, 4+ and 5+ for the Second Positive, Fourth Positive and Fifth Positive (Van der Ziel) Systems.

K.1 and *K*.2 for Kaplan's First and Second Systems.

The impulse-energy tensor of material particles

By T. S. Chang, *Department of Physics, National Central University, Chungking, China*

(*Communicated by P A. M. Dirac, F R.S.—Received 7 June* 1943)

PART I. MESONS AND ELECTRONS

The following is a direct and *general* construction, from the field quantities of the electrons and mesons and their derivatives, of a real, symmetrical and gauge-invariant tensor T_{kl} whose divergence to l is -1 times the four-force f_k experienced by the matter. Interpreting T_{kl} as the impulse-energy tensor of the material particles, for the case of no electromagnetic field the energy-momentum density obtained from T_{k4} is compared with that obtained from treating $(\hbar/i)(\partial/\partial x_k)$ as the energy-momentum operator Further, in the general case the Hamiltonian of the matter is compared with $\int T_{44}\,dx\,dy\,dz$. Both times, the two quantities compared agree apart from small modifications.

Introduction

As is well known, the conservation of the combined energy and momentum of matter and the electromagnetic field follows if (i) there exists a tensor T_{kl} satisfying

$$\frac{\partial}{\partial x_l} T_{kl} + f_k = 0, \tag{1}$$

f_k being the four-force experienced by the matter and (ii) T_{kl} is interpreted as the impulse-energy tensor belonging to matter. For a Dirac electron, Tetrode (1928) has found a symmetrical T_{kl} satisfying (1), which was later expressed in spinor form by Laporte & Uhlenbeck (1932). The search for T_{kl} was nearly completed by Heisenberg & Pauli (1929) who found in a general way an expression for T_{kl} satisfying (1) for all sorts of particles whose wave equations are deducible from varying a Lagrangian, but their expression is not symmetrical between k and l.

In Part I of the present paper, an attempt is made to obtain the most *general* symmetrical T_{kl} belonging to scalar and vector mesons and electrons. § I deals with scalar mesons interacting with electrons and neutrinos and § II with vector mesons interacting with the same light particles. The cases in which we have a single kind of particle only are obviously special cases. Owing to the significance of T_{kl} as the impulse-energy tensor of matter, such an attempt is certainly worth while.

Two things are to be observed regarding this impulse-energy tensor of matter. Let us take the case of electrons for definiteness. First, in the case of no electromagnetic field, where a possible solution for the wave functions of the electrons contains x_r only through the factor $\exp(ic_r x_r)$ (c_r being a constant four-vector), the

usual concept of treating $(\hbar/i)(\partial/\partial x_r)$ as the momentum operator leads one to expect $\rho c_r \hbar$ to be the energy-momentum density (ρ being the probability density of electrons) and thus to be equal to $(i/c)T_{k4}$. That this is nearly the case will be shown later. Secondly, one expects the integral of T_{44} over all space to be that of the Hamiltonian H for matter, which has the property that, from the rules of quantizing the field quantities ψ and from equations of the type

$$\frac{\partial \psi}{\partial t} = \frac{i}{\hbar}\left[\int H\,dV, \psi\right], \quad dV = dx\,dy\,dz,$$

one can obtain the wave equations for ψ. This is also found to be true apart from an unimportant modification. These results are certainly satisfactory.

I. Scalar mesons and electrons

Let x_r be the four-vector x, y, z, ict, ϕ_r be the electromagnetic four-potential A_x, A_y, A_z, $i\phi$ (A_x, A_y, A_z being the vector potential and ϕ the scalar potential), and Π_l, Π_l^*† be

$$-i\frac{\partial}{\partial x_l} - \frac{e}{\hbar c}\phi_l$$

and

$$i\frac{\partial}{\partial x_l} - \frac{e}{\hbar c}\phi_l \tag{2}$$

respectively. Before writing out the wave equations for A, A^* (the wave functions of scalar mesons), χ_μ^E, ψ_ν^E, χ_β^{E*}, ψ_ν^{E*} (those for electrons) and χ_μ^N, ψ_ν^N, χ_β^{N*}, ψ_ν^{N*} (those for neutrinos), their Lagrangians will be written out. With δ_l standing for $\partial/\partial x_l$ (which, it is assumed, does not operate through a bracket), $\delta_{\mu\beta}$ for the spinor of δ_l, $\Pi_{\mu\beta}$ for the spinor of Π_l, etc., and with κ^S, κ^E and κ^N as three constants connected

† In this paper, a tensor with a * symbol such as $T^*_{ik\ldots}$ is defined by

$$T^*_{ik\ldots} = (-1)^n \overline{(T_{ik\ldots})},$$

where n is the number of suffices among i, k, . . . which equal 4 and $\bar{\gamma}$ the complex conjugate of γ. (It is understood that by 'complex conjugate' is not meant numerically complex conjugate. This is evident since $\overline{T_{ik\ldots}}$ and $T_{ik\ldots}$ are different kinds of q-numbers.) For lack of better names, the tensor $T^*_{ik\ldots}$ is also called the complex conjugate of $T_{ik\ldots}$, and thus the complex conjugate of $T_{ik\ldots}$ may mean either $T^*_{ik\ldots}$ or $\overline{T_{ik\ldots}}$. For a spinor such as $t_{m\lambda\ \dot{n}\ldots}$, one can introduce a complex conjugate spinor $t^*_{m\lambda\ldots}$ defined by

$$t^*_{n\ldots m\lambda\ldots} = \overline{(t_{m\lambda\ldots n\ldots})},$$

where $\bar{\gamma}$ denotes the complex conjugate of γ as before. When t has an equal number of dotted and undotted suffices, the tensors corresponding to t and t^* are conjugate tensors. In the following the complex conjugate of a component $t_{m\lambda\ \dot{n}\ldots}$ of a spinor t with an equal number of dotted and undotted suffices may mean $t^*_{m\lambda\ \dot{n}\ldots}$ instead of $\overline{t_{m\lambda\ \dot{n}\ldots}}$. It is hoped that no confusion will arise from this double usage of the term 'complex conjugate'.

to the mass of scalar mesons, electrons and neutrinos by $\kappa^S\hbar = m^Sc$, $\kappa^E\hbar = m^Ec$ and $\kappa^N\hbar = m^Nc$, the Lagrangians are†

$$L^E = \hbar c[\kappa^E(\psi^{E\lambda}\chi^E_\lambda - \psi^E_\sigma\chi^{E\sigma}) + \tfrac{1}{2}\chi^{E\dot\mu}\Pi_{\lambda\dot\mu}\chi^{E\lambda} + \tfrac{1}{2}\chi^{E\lambda}\Pi^*_{\lambda\dot\mu}\chi^{E\dot\mu} + \tfrac{1}{2}\psi^{E\lambda}\Pi_{\lambda\dot\mu}\psi^{E\dot\mu} + \tfrac{1}{2}\psi^{E\dot\mu}\Pi^*_{\lambda\dot\mu}\psi^{E\lambda}], \tag{3·1}$$

$$L^N = \hbar c[\kappa^N(\psi^{N\lambda}\chi^N_\lambda - \psi^N_\sigma\chi^{N\sigma}) - \tfrac{1}{2}i\chi^{N\dot\mu}\delta_{\lambda\dot\mu}\chi^{N\lambda} + \tfrac{1}{2}i\chi^{N\lambda}\delta_{\lambda\dot\mu}\chi^{N\dot\mu} - \tfrac{1}{2}i\psi^{N\lambda}\delta_{\lambda\dot\mu}\psi^{N\dot\mu} + \tfrac{1}{2}i\psi^{N\dot\mu}\delta_{\lambda\dot\mu}\psi^{N\lambda}], \tag{3·2}$$

$$L^S = -\frac{\hbar c}{\kappa^S}\{(\kappa^S)^2(AA^* + F_lF_l^*) - g_2g_2^*M_lM_l^* - \kappa^S(g_1NA^* + g_1^*N^*A)\}, \tag{3·3}$$

where

$$\left.\begin{aligned} \kappa^SF_l &= \imath\Pi_lA + g_2M_l, \\ M_l &= \sigma_{l\dot\mu\dot\nu}(\psi^{N\dot\mu}\psi^{E\dot\nu} + l_1\chi^{E\dot\mu}\chi^{N\dot\nu}), \\ N &= l_2\psi^{N\dot\mu}\chi^E_{\dot\mu} + \chi^{N\dot\mu}\psi^E_{\dot\mu}, \end{aligned}\right\} \tag{4}$$

the $\sigma_{l\dot\mu\dot\nu}$'s being the usual Pauli matrices and g_1, g_2, l_1, l_2 being arbitrary complex constants.‡ Varying $\int(L^S + L^E + L^N)\,dV\,dt$ to A^*, $\psi^{\lambda E}$, $\chi^{\sigma E}$, $\psi^{\sigma N}$, $\chi^{\lambda N}$, one finds

$$-i\Pi_lF_l + \kappa^SA - g_1N = 0, \tag{5·1}$$

$$\kappa^E\chi^E_\lambda + \Pi^*_{\dot\mu\lambda}\psi^{E\dot\mu} + \frac{g_2i}{\kappa^S}(\Pi^*_{\dot\mu\lambda}A^*)\psi^{N\dot\mu} - g_1A^*\chi^N_\lambda = 0, \tag{5·2}$$

$$\kappa^E\psi^E_\sigma - \Pi^*_{\sigma\dot\nu}\chi^{E\dot\nu} - \frac{g_2l_1i}{\kappa^S}(\Pi^*_{\sigma\dot\nu}A^*)\chi^{N\dot\nu} + g_1l_2A^*\psi^N_\sigma = 0, \tag{5·3}$$

$$\kappa^N\chi^N_\sigma - i\delta_{\sigma\dot\lambda}\psi^{N\dot\lambda} + \frac{g_2i}{\kappa^S}(\Pi^*_{\sigma\dot\nu}A^*)\psi^{E\dot\nu} + g_1l_2A^*\chi^E_\sigma = 0, \tag{5·4}$$

$$\kappa^N\psi^N_\lambda + i\delta_{\mu\dot\lambda}\chi^{N\dot\mu} - \frac{g_2l_1i}{\kappa^S}(\Pi^*_{\dot\mu\lambda}A^*)\chi^{N\dot\mu} - g_1A^*\psi^E_\lambda = 0. \tag{5·5}$$

For convenience of future description, the left-hand sides of the above equations will be referred to as $h(A)$, $h(\chi^E_\lambda)$, $h(\psi^E_\sigma)$, $h(\chi^N_\sigma)$ and $h(\psi^N_\lambda)$. If every term in the above equations is replaced by its complex conjugate, then equations whose left-hand sides will be referred to as $h(A^*)$, $h(\chi^E_\lambda)$, $h(\psi^E_\sigma)$, $h(\chi^N_\sigma)$ and $h(\psi^N_\lambda)$ are obtained. It is hoped that the common usage of the same notation 'h' to denote the left-hand sides of the different wave equations will not lead to misunderstandings.

Varying the combined Lagrangian of the matter and the electromagnetic field subject to $\delta_l\phi_l = 0$, one finds

$$\delta_i\delta_l\phi_l = -(4\pi/c)j_l,$$

where

$$j_l = c(\partial/\partial\phi_l)(L^S + L^E + L^N) = ec[iAF_l^* - iA^*F_l - \sigma_{l\dot\mu\dot\nu}(\chi^{E\dot\mu}\chi^{E\dot\nu} + \psi^{E\dot\mu}\psi^{E\dot\nu})] \tag{6}$$

† The *'s are dropped from $\chi^{E*}_{\dot\lambda}$, $\psi^{E*}_{\dot\nu}$, $\chi^{N*}_{\dot\lambda}$, $\psi^{N*}_{\dot\nu}$, since this will not cause confusion.

‡ In current literature g_1 and g_2 are real and l_1 and l_2 are unity, which is less general.

and satisfies $\delta_l j_l = 0$ on account of equations (5·1–5·5) The problem now is to find T_{kl} satisfying

$$\delta_l T_{kl} + \frac{1}{c}(\delta_k \phi_l - \delta_l \phi_k) j_l = 0, \tag{7}$$

when the right-hand side of (6) is substituted for j_l.

Now impose the following conditions on T_{kl}:

(i) that it is gauge-invariant,

(ii) that the terms in T_{kl} involving A and A^* only do not contain symbolically more than two δ's and the terms involving χ^E_μ, ψ^E_ν, χ^E_μ and ψ^E_ν only (or χ^N_μ, ψ^N_ν, χ^N_μ, ψ^N_ν only) do not contain symbolically more than one δ, and

(iii) that T_{kl} is real and symmetrical between k and l. T_{kl} can now be written down as a sum of terms such as

$$A^* \Pi_k \Pi_l A, \quad \delta_{kl} \Pi^*_s A^* \Pi_s A, \quad \sigma^{m\dot{n}}_k \sigma^{\lambda\dot{s}}_l \psi^E_{\dot{n}} \Pi^*_{\lambda\dot{s}} \psi^E_m, \tag{8}$$

etc., each with an unknown coefficient γ Now split the coefficients of all terms containing A and A^* only into some power of κ^S times some quantities so that the latter are of the same dimensions as the coefficient of $A^* \Pi_k \Pi_l A$. It is at once found that the terms with some power of κ^S appearing in the coefficient can be eliminated with the help of the wave equations. Then assume that this elimination is carried out so that the coefficients of different terms containing A and A^* only contain no κ^S as a factor and are of the same dimensions. Let similar steps be taken toward the terms containing the wave functions of electrons (or neutrinos) only. After carrying out such manipulations, our assumed form for T_{kl} is substituted into (7), giving a long equation (8) which it is hoped can be made identically zero by properly choosing the value of the constants γ The different terms in (8) are not all independent; relations between them will be given below. Following the method of unknown multipliers, multiply the relations by unknown constants λ, add them to (8), set the coefficients of different terms in the resulting equation to zero and solve concurrently for the λ's and the γ's If in so doing no relation between terms in (8) has been left out, the result of solving for λ and γ gives the most general T_{kl} satisfying our conditions. All the algebra will be left out except for writing out for future reference the various relations between terms in (8) with their proper multiplying constants,

$$\tfrac{1}{2}\hbar c \quad \{-iA \quad \Pi^*_k \; h(A^*) + i(\Pi^*_k \; A^*) \quad h(A)\} \quad = 0, \tag{9}$$

$$\tfrac{1}{8}\hbar c \sigma^{m\dot{n}}_k \{ \quad i\psi^{E\rho} \Pi_{\rho\dot{n}} \, h(\chi^E_m) \quad - i(\Pi_{\rho\dot{n}} \chi^E_m) \quad h(\psi^{E\rho})\} = 0, \tag{10·1}$$

$$\tfrac{3}{8}\hbar c \sigma^{m\dot{n}}_k \{ \quad i\psi^E_m \; \Pi^\rho{}_{\dot{n}} h(\chi^E_\rho) \quad - i(\Pi^\rho{}_{\dot{n}} \chi^E_\rho) \quad h(\psi^E_m)\} \quad = 0, \tag{10·2}$$

$$\tfrac{3}{8}\hbar c \sigma^{m\dot{n}}_k \{ -i\chi^E_m \; \Pi^*_{\rho\dot{n}} h(\psi^{E\rho}) + i(\Pi^*_{\rho\dot{n}} \psi^{E\rho}) \, h(\chi^E_m)\} \quad = 0, \tag{10·3}$$

$$\tfrac{1}{8}\hbar c \sigma^{m\dot{n}}_k \{ -i\chi^E_\rho \; \Pi^{*\rho}_{\dot{n}} h(\psi^E_m) + i(\Pi^{*\rho}_{\dot{n}} \psi^E_m) \; h(\chi^E_\rho)\} \quad = 0, \tag{10·4}$$

$$\tfrac{1}{8}\hbar c \sigma^{m\dot{n}}_k \{ \quad \psi^{N\rho} \delta_{\rho\dot{n}} \; h(\chi^N_m) \quad - (\delta_{\rho\dot{n}} \; \chi^N_m) \quad h(\psi^{N\rho})\} = 0, \tag{11·1}$$

$$\tfrac{3}{8}\hbar c \sigma^{m\dot{n}}_k \{ \quad \psi^N_m \delta^\rho{}_{\dot{n}} \; h(\chi^N_\rho) \quad - (\delta^\rho{}_{\dot{n}} \; \chi^N_\rho) \quad h(\psi^N_m)\} \quad = 0, \tag{11·2}$$

$$\tfrac{3}{8}\hbar c \sigma^{m\dot{n}}_k \{ \quad \chi^N_m \delta_{\rho\dot{n}} \; h(\psi^{N\rho}) \quad - (\delta_{\rho\dot{n}} \; \psi^{N\rho}) \, h(\chi^N_m)\} \quad = 0, \tag{11·3}$$

$$\tfrac{1}{8}\hbar c \sigma^{m\dot{n}}_k \{ \quad \chi^N{}_\rho \delta^\rho{}_{\dot{n}} \; h(\psi^N_m) \quad - (\delta^\rho{}_{\dot{n}} \; \psi^N_m) \; h(\chi^N_\rho)\} \quad = 0, \tag{11·4}$$

together with the conjugate complex equations.

The replacement of Π and Π^* by $-i\delta$ and $i\delta$ in equations (10) and the reverse replacement in (11) give in fact a number of other relations, but their multiplying constants are all zero.†

In giving the result for T_{kl}, it is convenient to put it as

$$Y_{kl}+Y_{lk}+Y^*_{kl}+Y^*_{lk}, \tag{12}$$

and write out Y_{kl}. It falls into four parts:

(i) The part that concerns A and A^* only,

$$\gamma^{(1)}\{\Pi^*_k A^*\Pi_l A - A^*\Pi_k\Pi_l A + \delta_{kl}A^*\Pi_s\Pi_s A - \delta_{kl}\Pi^*_s A^*\Pi_s A\} \\ -\tfrac{1}{4}(\hbar c/\kappa^S)\{\Pi^*_k A^*\Pi_l A + A^*\Pi_k\Pi_l A\}, \tag{13}$$

where $\gamma^{(1)}$ is an arbitrary real constant. This will be denoted by Y^S_{kl} and the corresponding T_{kl} by T^S_{kl}.

(ii) The part that concerns electrons only,

$$\tfrac{1}{4}\sigma^{mn}_k\sigma^{\lambda\delta}_l\{\gamma^{(2)}y^{(2)}_{m\lambda n\delta}+\gamma^{(3)}y^{(3)}_{m\lambda n\delta}+\ldots+\gamma^{(7)}y^{(7)}_{m\lambda n\delta}+y^{(8)}_{m\lambda n\delta}\}, \tag{14}$$

where $\gamma^{(2)}$, $\gamma^{(3)}$, . are arbitrary real constants,

$$y^{(2)}_{m\lambda n\delta} = \psi^E_n\Pi^*_{m\delta}\psi^E_\lambda - \psi^E_n\Pi^*_{\lambda\delta}\psi^E_m + \tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{n\delta}\psi^E_\rho\Pi^{*\rho\theta}\psi^E_\theta, \tag{15·1}$$

$$y^{(3)}_{m\lambda n\delta} = -\epsilon_{m\lambda}\psi^E_\delta\Pi^{*\alpha}_n\psi^E_\alpha + \tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{n\delta}\psi^E_\rho\Pi^{*\rho\theta}\psi^E_\theta. \tag{15·2}$$

$$y^{(4)}_{m\lambda n\delta} = -\epsilon_{n\delta}\psi^E_\rho\Pi^{*\rho}_m\psi^E_\lambda + \tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{n\delta}\psi^E_\rho\Pi^{*\rho\theta}\psi^E_\theta, \tag{15·3}$$

$$y^{(8)}_{m\lambda n\delta} = -\tfrac{1}{2}\hbar c(\psi^E_n\Pi^*_{\lambda\delta}\psi^E_m + \chi^E_n\Pi_{\lambda\delta}\chi^E_m), \tag{15·4}$$

($\epsilon_{m\lambda}$, $\epsilon_{n\delta}$ being the well-known antisymmetrical spinors), and the $y^{(5)}$, $y^{(6)}$, $y^{(7)}$ are spinors obtained from $y^{(2)}$, $y^{(3)}$, $y^{(4)}$ by replacing ψ by χ and Π^* by Π with all suffices remaining unchanged. This Y will be denoted by Y^E, and the corresponding T by T^E

(iii) The part that concerns neutrinos only, this can be obtained from (ii) by putting e to be zero. This Y will be denoted by Y^N, and its T by T^N.

(iv) The part that contains the g's,

$$\tfrac{1}{2}\hbar c\left\{\tfrac{1}{2}\delta_{kl}g_1A^*N + \tfrac{1}{2}\delta_{kl}\frac{g_2 i}{\kappa^S}A^*\Pi_s M_s + \frac{g_2 i}{\kappa^S}M_l\Pi^*_k A^*\right\}. \tag{16}$$

This will be denoted by Y^I (the index 'I' signifying interaction), and its T by T^I. This completes the answer for T_{kl}.

† Strictly speaking, the whole set of relations between terms in (8) is not completed yet. By permuting in a proper way the spinor suffices m, n, etc., in (10) and (11) (either with Π, Π^* replaced by $-i\delta$, $i\delta$ or not), new equations are derived, but these can also be obtained from the already existing ones by employing the rules

$$a_\lambda b^\lambda = -a^\lambda b_\lambda, \quad a_\lambda b^\lambda c_m + a_m b_\lambda c^\lambda + a^\lambda b_m c_\lambda = 0,$$

and are hence not independent equations. Further, there are equations with one side zero and the other side $(\delta_l\phi_l)$ times some expressions, which obviously may form part of the relations between terms in (8). Relations of such kinds will not be written out here.

Now see what happens to T_{kl} for the special case in which $\phi = 0$, χ_μ^E and ψ_μ^E contain x_r through the factor $\exp(ic_r^E x_r)$, χ_μ^N and ψ_μ^N through the factor $\exp(ic_r^N x_r)$ and A through the factor $\exp(ic_r^S x_r)$. In the first place

$$c_r^S = c_r^E - c_r^N, \quad (\Pi_l A)\, M_l^* = A\Pi_l^* M_l^*,$$

etc For electrons, on putting all the γ's in (14) zero, it is found that

$$T_{kl}^E = -\frac{\hbar c}{4}\sigma_k^{m\dot{n}}\sigma_l^{\lambda\dot{s}}\{c_{m\dot{n}}^E(\chi_\lambda^E\chi_{\dot{s}}^E + \psi_\lambda^E\psi_{\dot{s}}^E) + c_{\lambda\dot{s}}^E(\chi_m^E\chi_{\dot{n}}^E + \psi_m^E\psi_{\dot{n}}^E)\}. \tag{17}$$

Now
$$\begin{aligned}\sigma_l^{\lambda\dot{s}}c_{m\dot{n}}^E\chi_\lambda^E\chi_{\dot{s}}^E &= -\sigma_{lm}^\lambda c_{m\dot{s}}^E\chi^{E\dot{s}}\chi_\lambda^E - \sigma_{l\dot{n}}^\lambda c_m^{E\dot{s}}\chi_{\dot{n}}^E\chi_\lambda^E \\ &= -\sigma_{lm}^\lambda c_{m\dot{s}}^E\chi^{E\dot{s}}\chi_\lambda^E + \sigma_{lm\dot{s}}c_\lambda^{E\dot{s}}\chi_{\dot{n}}^E\chi^{E\lambda} + \sigma_{l\lambda\dot{s}}c^{E\lambda\dot{s}}\chi_{\dot{n}}^E\chi_m^E.\end{aligned}$$

Hence T_{kl}^E is the sum of two parts, one with its T_{k4} equal to $-i\hbar c c_k^E\rho^E$ (ρ^E being the probability density of the electrons $\sum_{\lambda=1,2}(\chi_\lambda^E\chi_\lambda^E + \psi_\lambda^E\psi_\lambda^E)$) and the other given by

$$-\frac{\hbar c}{4}\sigma_k^{m\dot{n}}\{\sigma_{lm}^\lambda c_{m\dot{s}}^E(\chi^{E\dot{s}}\chi_\lambda^E + \psi^{E\dot{s}}\psi_\lambda^E) - \sigma_{lm\dot{s}}c_\lambda^{E\dot{s}}(\chi_{\dot{n}}^E\chi^{E\lambda} + \psi_{\dot{n}}^E\psi^{E\lambda})\}. \tag{18}$$

Similarly, T_{kl}^N is the sum of two parts, one with its T_{k4} equal to $-i\hbar c c_k^N\rho^N$ and the other given by something similar to (18) Eliminating from the latter and (18) $c_{m\dot{s}}^E\chi^{E\dot{s}}$, $c_{m\dot{s}}^E\psi^{E\dot{s}}$, $c_\lambda^{E\dot{s}}\chi^{E\lambda}$, $c_\lambda^{E\dot{s}}\psi^{E\lambda}$, etc., by means of the wave equations, so that the resulting expressions contain no other derivatives than that of A, and noting that each must be antisymmetrical between k and l, it is found that their sum reduces to

$$\frac{\hbar c}{2}\frac{g_2\imath}{\kappa^S}\{-M_k\Pi_l^*A^* + M_l\Pi_k^*A^*\} + \text{comp. conj}$$

Finally, for any value of $\gamma^{(1)}$ in (13), it is observed that

$$T_{k4}^S + \frac{\hbar c}{\kappa^S}(g_2 i M_4\Pi_k^*A^* - g_2^* i M_4^*\Pi_k A) = -i\hbar c c_k^S\rho^S,$$

ρ^S being the probability density of the scalar mesons, and thus

$$T_{kl} = -i\hbar c(c_k^S\rho^S + c_k^E\rho^E + c_k^N\rho^N) - \tfrac{1}{2}\hbar c\Delta\delta_{kl}, \tag{19}$$

where Δ is short for

$$-g_1A^*N - (g_2\imath/\kappa^S)\,M_l\Pi_l^*A^* + \text{comp. conj.} \tag{20·1}$$

which in the present special case $\phi = 0$ reduces to

$$2\kappa^E(\chi_\lambda^E\psi^{E\lambda} + \chi_{\dot{s}}^E\psi^{E\dot{s}}) + 2c_{\lambda\dot{s}}^E(\chi^{E\lambda}\chi^{E\dot{s}} + \psi^{E\lambda}\psi^{E\dot{s}}) \tag{20·2}$$

or
$$2\kappa^N(\chi_\lambda^N\psi^{N\lambda} + \chi_{\dot{s}}^N\psi^{N\dot{s}}) + 2c_{\lambda\dot{s}}^N(\chi^{N\lambda}\chi^{N\dot{s}} + \psi^{N\lambda}\psi^{N\dot{s}}) \tag{20·3}$$

or
$$-2(\kappa^S)^{-1}[(\kappa^S)^2 + c_r^Sc_r^S]\,AA^*. \tag{20·4}$$

Thus apart from the term $-\frac{1}{2}\hbar c\Delta\delta_{kl}$ in (19), $(i/c)\,T_{kl}$ gives the expected value of momentum density.

It is impossible to compare T_{44} with the Hamiltonian H of matter, since the former is gauge-invariant while the latter is not. Let P be the canonical variable to A, P^* that to A^*, and let H' be

$$H+\frac{e}{\hbar}\phi_4(AP-A^*P^*)+ie\phi_4(\chi^{E\lambda}\chi^{E\lambda}+\psi^{E\mu}\psi^{E\mu}), \tag{21}$$

so that the wave equations take the gauge-invariant form

$$\left.\begin{aligned}
\Pi_4 A &= -\frac{i}{\hbar c}\Big[\int H'dV, A\Big], & \Pi_4^* P &= \frac{i}{\hbar c}\Big[\int H'dV, P\Big],\\
\Pi_4^* A^* &= \frac{i}{\hbar c}\Big[\int H'dV, A^*\Big], & \Pi_4 P^* &= -\frac{i}{\hbar c}\Big[\int H'dV, P^*\Big],\\
\Pi_4 \psi^{E\dot\mu} &= -\frac{i}{\hbar c}\Big[\int H'dV, \psi^{E\dot\mu}\Big], & \Pi_4^* \psi^{E\mu} &= \frac{i}{\hbar c}\Big[\int H'dV, \psi^{E\mu}\Big],\\
\frac{\partial}{\partial t}\psi^{N\dot\mu} &= \frac{i}{\hbar}\Big[\int H'dV, \psi^{N\dot\mu}\Big], & \frac{\partial}{\partial t}\psi^{N\mu} &= \frac{i}{\hbar}\Big[\int H'dV, \psi^{N\mu}\Big],
\end{aligned}\right\} \tag{22}$$

etc. Then, adding to the above T_{kl} the term $-\delta_{kl}$ times

$$L^E+L^N-\hbar c\Lambda-\frac{\hbar c}{2\kappa^S}\{[(\kappa^S)^2 AA^*+A^*(\Pi_l\Pi_l A-g_1\kappa^S N-g_2\iota\Pi_l M_l)]+\text{comp. conj.}\}, \tag{23}$$

which vanishes always, then by partial integrations over the space co-ordinates

$$\int H'dV=\int T_{44}dV, \tag{24}$$

showing that the modified Hamiltonian $\int H'dV$ is the total energy.

II. Vector mesons and electrons

The Lagrangian of the vector mesons is

$$-\frac{\hbar c}{\kappa^V}\{(\kappa^V)^2 A_i A_i^*+\tfrac{1}{2}(\kappa^V)^2 F_{ij}F_{ij}^*-\tfrac{1}{2}g_2 g_2^*\Gamma_{ij}\Gamma_{ij}^*-\kappa^V(g_1 M_i A_i^*+g_1^* M_i^* A_i)\}, \tag{25}$$

where

$$\kappa^V F_{ij}=i(\Pi_i A_j-\Pi_j A_i)+g_2\Gamma_{ij}, \tag{26·1}$$

$$M_l=\sigma_{l\mu\dot\nu}(\psi^{N\mu}\psi^{E\dot\nu}+l_1\chi^{N\dot\nu}\chi^{E\mu}), \tag{26·2}$$

$$\Gamma_{ij}=\sigma_{i,n\dot\alpha}\sigma_{j,m\dot\beta}[l_2\epsilon^{\dot\alpha\dot\beta}(\psi^{Nn}\chi^{Em}+\psi^{Nm}\chi^{En})+\epsilon^{nm}(\psi^{E\dot\alpha}\chi^{N\dot\beta}+\psi^{E\dot\beta}\chi^{N\dot\alpha})], \tag{26·3}$$

g_1, g_2, l_1, l_2 being again arbitrary real or complex constants. With L^E and L^N given in the previous section, it is easy to see by performing variations that

$$\kappa^V A_j-i\Pi_i F_{ij}-g_1 M_j=0, \tag{27·1}$$

$$\kappa^E \chi^E_\lambda + \Pi^*_{\mu\lambda}\psi^{E\mu} + g_1 A^*_{\nu\lambda}\psi^{N\nu} + \frac{g_2 i}{\kappa^V}(\Pi^*_{\alpha\lambda}A^{*\alpha}_{\mu} - \Pi^{*\alpha}_{\mu}A^*_{\alpha\lambda})\chi^{N\mu} = 0, \qquad (27{\cdot}2)$$

$$\kappa^E \psi^E_\sigma - \Pi^*_{\sigma\lambda}\chi^{E\lambda} - g_1 l_1 A^*_{\sigma\mu}\chi^{N\mu} - \frac{g_2 l_3 i}{\kappa^V}(\Pi^*_{\mu\alpha}A^{*\dot{\alpha}}_{\dot{\sigma}} - \Pi^{*\dot{\alpha}}_{\dot{\sigma}}A^*_{\mu\alpha})\psi^{N\mu} = 0, \qquad (27{\cdot}3)$$

$$\kappa^N \chi^N_\sigma - i\delta_{\sigma\lambda}\psi^{N\lambda} + g_1 A^*_{\sigma\mu}\psi^{E\mu} + \frac{g_2 l_3 i}{\kappa^V}(\Pi^*_{\sigma\dot{\alpha}}A^{*\dot{\alpha}}_{\mu} - \Pi^{*\alpha}_{\mu}A^*_{\sigma\dot{\alpha}})\chi^{E\mu} = 0, \qquad (27{\cdot}4)$$

$$\kappa^N \psi^N_\lambda + i\delta_{\mu\lambda}\chi^{N\mu} - g_1 l_1 A^*_{\sigma\lambda}\chi^{E\sigma} - \frac{g_2 i}{\kappa^V}(\Pi^*_{\alpha\dot{\beta}}A^{*\alpha}_{\lambda} - \Pi^{*\alpha}_{\lambda}A^*_{\alpha\dot{\beta}})\psi^{E\dot{\beta}} = 0. \qquad (27{\cdot}5)$$

As before, the left-hand sides of the above equations and their conjugates will be referred to as $h(A_j)$ or $h(\chi^E_\lambda)$, $h(\psi^E_\sigma)$, $h(\chi^N_\sigma)$, etc Varying the combined Lagrangian of matter and field, one finds

$$\delta_l \delta_\iota \phi_l = -\frac{4\pi}{c} j_l,$$

where j_l is given by

$$c(\partial/\partial\phi_l)(L^V + L^E + L^N) = ec\{iA_i F^*_{li} - iA^*_i F_{li} - \sigma_{\iota,\mu\dot{\nu}}(\psi^{E\mu}\psi^{E\dot{\nu}} + \chi^{E\mu}\chi^{E\dot{\nu}})\}, \qquad (28)$$

and satisfies $\delta_l j_l = 0$ on account of the wave equations

The task is to substitute this j_l into

$$\delta_l T_{kl} + \frac{1}{c}(\delta_k\phi_l - \delta_l\phi_k)j_l = 0$$

and find the corresponding T_{kl}. The same set of conditions will be imposed on T_{kl} as before and the same procedure employed in finding it The only difference is that equation (9) is to be replaced by

$$\left.\begin{aligned} \tfrac{1}{2}\hbar c\{-\iota A^*_k \Pi_\iota h(A_i) + \iota(\Pi_\iota A_i)h(A^*_k)\} &= 0, \\ \tfrac{1}{2}\hbar c\{-iA^*_i \Pi_k h(A_i) + \iota(\Pi_k A_i)h(A^*_i)\} &= 0, \\ \tfrac{1}{2}\hbar c\{\quad \iota A^*_i \Pi_\iota h(A_k) - \iota(\Pi_\iota A_k)h(A^*_i)\} &= 0, \end{aligned}\right\} \qquad (29)$$

and that the symbols h now denote entirely different functions. Writing T_{kl} as $Y_{kl} + Y_{lk} + Y^*_{kl} + Y^*_{lk}$, the result for Y_{kl} falls into four parts, of which that concerning electrons and that concerning neutrinos are exactly those given in the previous section. The other two parts are·

(i) That concerning vector mesons only,

$$\begin{aligned} &\gamma^{(8)}\{\Pi^*_s A^*_k \Pi_s A_l - \Pi^*_k A^*_s \Pi_\iota A_l - \Pi^*_k A^*_l \Pi_s A_s - A^*_k \Pi_s \Pi_s A_l \\ &\quad + A^*_k \Pi_l \Pi_s A_s + A^*_s \Pi_k \Pi_s A + \tfrac{1}{2}\delta_{kl}\Pi^*_r A^*_s \Pi_s A_r \\ &\quad + \tfrac{1}{2}\delta_{kl}\Pi^*_r A^*_r \Pi_s A_s - \tfrac{1}{2}\delta_{kl}A^*_r \Pi_r \Pi_s A_s - \tfrac{1}{2}\delta_{kl}A^*_r \Pi_s \Pi_r A_s\} \\ &+ \gamma^{(9)}\{\Pi^*_k A^*_s \Pi_l A_s - A^*_s \Pi_k \Pi_l A_\iota + \delta_{kl}A^*_r \Pi_s \Pi_s A_r - \delta_{kl}\Pi^*_r A^*_s \Pi_r A_s\} \\ &- \frac{\hbar c}{2\kappa^V}\{\Pi^*_k A^*_l \Pi_\iota A_s + \Pi^*_k A^*_s \Pi_l A_s - \Pi^*_k A^*_s \Pi_s A_l \\ &\quad - A^*_s \Pi_s \Pi_l A_k - \tfrac{1}{2}\delta_{kl}\Pi^*_r A^*_s \Pi_r A_s - \tfrac{1}{2}\delta_{kl}\Pi^*_r A^*_r \Pi_s A_s \\ &\quad + \tfrac{1}{2}\delta_{kl}A^*_r \Pi_r \Pi_s A_s + \tfrac{1}{2}\delta_{kl}A^*_r \Pi_s \Pi_s A_r\}, \end{aligned} \qquad (30)$$

where $\gamma^{(8)}$ and $\gamma^{(9)}$ are arbitrary real constants This is denoted by Y^V and its T by T^V.

(ii) That containing either g_1 or g_2,

$$-\tfrac{1}{2}\hbar c\frac{g_2 i}{\kappa^V}\{\Gamma_{sk}(\Pi_l^* A_s^* - \Pi_s^* A_l^*) + A_l^* \Pi_s \Gamma_{sk}\} + \tfrac{1}{2}\hbar c\,\delta_{kl}\left\{g_1 A_i^* M_i + \frac{g_2 i}{\kappa^V} A_i^* \Pi_s \Gamma_{si}\right\}. \quad (31)$$

This is denoted by Y^I and its T by T^I.

For the special case in which $\phi = 0$ and the solution for χ_μ^E, χ_μ^N and A_i contains x_r through the factors $\exp(ic_r^E x_r)$, $\exp(ic_r^N x_r)$ and $\exp(ic_r^V x_r)$, it is noted in the same way that after setting all the γ's in (14) zero, T_{kl}^E can be reduced to two parts, one with its T_{kk} equal to $-i\hbar c c_k^E \rho^E$ and the other given by (18). (18) and a similar expression arising from T_{kl}^N combine to give

$$\tfrac{1}{2}\hbar c\frac{g_2 i}{\kappa^V}[\Gamma_{ki}(\Pi_i^* A_l^* - \Pi_l^* A_i^*) + \Gamma_{li}(\Pi_k^* A_i^* - \Pi_i^* A_k^*)] + \tfrac{1}{2}\hbar c g_1(-A_l^* M_k + A_k^* M_l) + \text{comp. conj.}$$

Noting that

$$-\frac{g_2 i}{\kappa^V}(\Gamma_{li}\Pi_i^* A_k^* + \Gamma_{sk}\Pi_s^* A_l^*) + g_1(-A_l^* M_k + A_k^* M_l) - \frac{1}{\kappa^V}(\Pi_k A_s \Pi_s^* A_l^* - \Pi_l A_s \Pi_s^* A_k^*) + \text{comp. conj.} = 0$$

and that

$$-\frac{1}{\kappa^V}(c_4^V A_s A_s^* - c_s^V A_s A_4^*) + \frac{g_2 i}{\kappa^V}\Gamma_{4i} A_i^* + \text{comp. conj} = -i\rho^V,$$

where ρ^V is the probability density of the vector mesons, one obtains finally for any value of $\gamma^{(8)}$ and $\gamma^{(9)}$

$$T_{kk} = -i\hbar c(c_k^V \rho^V + c_k^E \rho^E + c_k^N \rho^N) - \tfrac{1}{2}\hbar c\Delta\,\delta_{kk}, \quad (32)$$

where

$$\Delta = -g_1 A_i^* M_i - \frac{g_2 i}{\kappa} A_l^* \Pi_s \Gamma_{si} + \text{comp. conj}, \quad (33{\cdot}1)$$

which, expressed in the wave functions of electrons (or neutrinos) for the present special case, reduces exactly to the expression (20·2) or (20·3), and expressed in the wave functions of the vector mesons, to

$$-\frac{1}{\kappa^V}[(\kappa^V)^2 A_i A_i^* + c_i^V c_i^V A_j A_j^* - c_i^V c_j^V A_i A_j^*] + \text{comp. conj.} \quad (33{\cdot}2)$$

Before concluding this section, consider the relation between the Hamiltonian of matter H and T_{44}. Let P_i be the canonical variable to A_i ($i = 1, 2, 3$) and P_i^* that to A_i^* ($i = 1, 2, 3$), and let

$$H' = H + \frac{e}{\hbar}\phi_4 \sum_{i=1}^{3}(P_i A_i - P_i^* A_i^*) + ie\phi_4(\chi^{E\lambda}\chi^{E\lambda} + \psi^{E\mu}\psi^{E\mu}). \quad (34)$$

$$\left.\begin{aligned} &\text{Then}\quad \Pi_4 A_i = -\frac{i}{\hbar c}\Big[\int H'dV, A_i\Big], \quad \Pi_4^* P_i = \frac{i}{\hbar c}\Big[\int H'dV, P_i\Big], \quad (i = 1,2,3)\\ &\Pi_4^* A_i^* = \frac{i}{\hbar c}\Big[\int H'dV, A_i^*\Big], \qquad \Pi_4 P_i^* = -\frac{i}{\hbar c}\Big[\int H'dV, P_i^*\Big], \end{aligned}\right\} \tag{35}$$

equivalent to the wave equations for A_i, A_i^*, P_i and P_i^* $(i = 1, 2, 3)$, and

$$\Pi_4 \psi^{EA} = -\frac{i}{\hbar c}\Big[\int H'dV, \psi^{EA}\Big],$$

etc., equivalent to the wave equations for ψ^{EA}, etc. It is satisfactory to see that if to the above T_{kl} is added $-\delta_{kl}$ times

$$L^E + L^N - \hbar c\Delta - \frac{\hbar c}{2\kappa^V}\{[(\kappa^V)^2 A_i A_i^* + A_i^*(\Pi_j \Pi_j A_i - \Pi_j \Pi_i A_j - g_2{}^2 \Pi_j \Gamma_{ji} - g_1 \kappa^V M_i)] + \text{comp. conj.}\}, \tag{36}$$

which vanishes always, by partial integrations over the space coordinates

$$\int H'dV = \int T_{44} dV$$

is obtained. Thus the Hamiltonian is again the total energy.

It is perhaps needless to add that the above calculations and conclusions can be extended to include protons and neutrons So long as there is no direct interaction between the heavy and the light particles, the extension can be easily made. For this reason, it will be left out here

PART II. PARTICLES OF SPIN 2 OR 3/2

Calculations of a general impulse-energy tensor T_{kl} for mesons and electrons are here extended to particles of spin 2 or 3/2, whose wave equations were given by Fierz & Pauli. Results are generally similar, though expressions for T_{kl} are now much more complicated

Introduction

In part I of this paper, it is pointed out that for vector or scalar mesons and electrons, there exists a real and symmetrical tensor T_{kl} satisfying

$$\delta_l T_{kl} + f_k = 0, \tag{1}$$

where f_k is the four-force and δ_l denotes differentiation to x_l, (x_l being the vector x, y, z, ict). On interpreting T_{k4} $(k = 1, 2, 3)$ as $-ic$ times the momentum density of matter, T_{44} as the energy density of matter, the conservation law for the total energy and momentum of the matter and the electromagnetic field in any region of space is obtained, account being taken for their flow across the surface bounding the region. Needless to say, in constructing expressions for the energy or momentum of matter, such a conservation must be constantly kept in sight and thus the correct construction must start from determining T_{kl} from (1), obtaining the density expressions from T_{kl} and performing a final integration over space.

In this part of the paper the investigation of T_{kl} will be extended to particles of spin 2 and 3/2. Expressions for T_{kl} for such particles in a force-free field have been given by Fierz (1939), and thus here attention will be concentrated on T_{kl} for particles in a general electromagnetic field. Throughout the entire calculations, the wave equations given by Fierz & Pauli (1939) for such particles will be employed.

I. Particles of spin 2

Following Fierz, let the wave functions be a symmetrical tensor A_{kl} together with a scalar C and obtain the wave equations by varying a Lagrangian $L^{(2)}$. They are

$$2\kappa^2 A_{ik} + 2\Pi_l \Pi_l A_{ik} - 2\{\Pi_i \Pi_s A_{sk} + \Pi_k \Pi_s A_{si}\} - \frac{ei}{\hbar c}(\phi_{il} A_{lk} + \phi_{lk} A_{il})$$
$$+ \tfrac{1}{2}(\Pi_i \Pi_k + \Pi_k \Pi_i) C - \tfrac{1}{2}\delta_{ik}(-2\Pi_r \Pi_s A_{rs} + \tfrac{1}{2}\Pi_l \Pi_l C) = 0, \quad (2{\cdot}1)$$
$$2\kappa^2 C + \Pi_l \Pi_l C - \tfrac{1}{3}\Pi_r \Pi_s A_{rs} = 0, \quad (2{\cdot}2)$$

where ϕ_{il} denotes $\delta_i\phi_l - \delta_l\phi_i$ and all the other symbols have the same meaning as in part I. As before, the left-hand sides of the above equations will be called $h(A_{ik})$ and $h(C)$ and those of the conjugate equations $h(A^*_{ik})$ and $h(C^*)$.

Fierz has not given an expression for the current, but one can be obtained from varying the Lagrangian to ϕ_l. Varying the combined Lagrangian of field and matter subject to $\delta_l\phi_l = 0$, one finds

$$\frac{1}{4\pi}\delta_i\delta_i\phi_l - \delta_l\Phi = -\frac{1}{c}j_l,$$

where Φ is a scalar to be determined later by $\delta_l\phi_l = 0$, and j_l the vector

$$c\frac{\partial L}{\partial \phi_l} = \frac{ec}{\kappa}\{A^*_{ik}\Pi_l A_{ik} - A^*_{ik}\Pi_i A_{lk} - A^*_{lk}\Pi_s A_{sk}$$
$$+ \tfrac{1}{2}A^*_{kl}\Pi_k C + \tfrac{1}{2}C^*\Pi_k A_{kl} - \tfrac{3}{8}C^*\Pi_l C + \text{comp. conj}\}. \quad (3)$$

It is easy to verify from the wave equations that $\delta_l j_l = 0$, thus we take it as the current vector. At the same time, since $\delta_l j_l = 0$, then $\delta_l\delta_l\Phi = 0$ or $\Phi = 0$, and thus

$$\delta_i\delta_i\phi_l = -\frac{4\pi}{c}j_l, \quad (4)$$

as usual. Incidentally, it may be pointed out that the condition $\delta_l j_l = 0$ is not sufficient to determine j_l. An example is provided by the vector

$$A^*_{ik}\Pi_i A_{lk} - A^*_{lk}\Pi_i A_{ik} + \text{comp. conj.},$$

whose divergence is zero under all conditions.

As before the method is to substitute the right side of (3) for j_l into

$$\delta_l T_{kl} + \frac{1}{c}(\delta_k\phi_l - \delta_l\phi_k) j_l = 0$$

and solve for T_{kl}. The procedure of finding T_{kl} is exactly the same as before, and its

description will therefore be omitted here. The equations playing the part of (9), (10) and (11) in part I are

$$\left.\begin{aligned}
\tfrac{1}{4}\frac{\hbar c}{\kappa}\{-\imath A_{il}\Pi_k^*\, h(A_{il}^*)+\imath(\Pi_k^* A_{il}^*)\,h(A_{il})\} &= 0,\\
-\tfrac{1}{2}\frac{\hbar c}{\kappa}\{-\imath A_{il}\Pi_l^*\, h(A_{ik}^*)+\imath(\Pi_l^* A_{ik}^*)\,h(A_{il})\} &= 0,\\
\tfrac{1}{2}\frac{\hbar c}{\kappa}\{-\imath A_{ik}\Pi_l^* h(A_{il}^*)+\imath(\Pi_l^* A_{il}^*)\,h(A_{ik})\} &= 0,\\
0\{-\imath A_{ik}\Pi_l^* h(C^*)+\imath(\Pi_l^* C^*)\,h(A_{ik})\} &= 0,\\
0\{-\imath C\Pi_l^*\;\; h(A_{ik}^*)+\imath(\Pi_l^* A_{ik}^*)\,h(C)\} &= 0,\\
-\tfrac{3}{16}\frac{\hbar c}{\kappa}\{-\imath C\Pi_k^*\;\; h(C^*)+\imath(\Pi_k^* C^*)\,h(C)\} &= 0,
\end{aligned}\right\} \tag{5}$$

and their conjugates. The result for T_{kl} is that, if it is written as

$$Y_{kl}+Y_{lk}+Y_{kl}^*+Y_{lk}^*,$$

Y_{kl} is given by

$$\gamma^{(1)}Y_{kl}^{(1)}+\gamma^{(2)}Y_{kl}^{(2)}+\ldots+\gamma^{(5)}Y^{(5)}+Y^{(6)}, \tag{6}$$

where $\gamma^{(1)}$, $\gamma^{(2)}$, . are arbitrary real constants and

$$\begin{aligned}
Y_{kl}^{(1)} &= \Pi_l^* A_{lr}^*\Pi_r A_{kl}+\Pi_s^* A_{ls}^*\Pi_r A_{kr}-2\Pi_r^* A_{rs}^*\Pi_s A_{kl}\\
&\quad -A_{kr}^*\Pi_s\Pi_r A_{ls}-A_{kr}^*\Pi_r\Pi_s A_{ls}+A_{rs}^*\Pi_r\Pi_s A_{kl}+A_{kl}^*\Pi_r\Pi_s A_{rs}, &(7\cdot1)\\
Y_{kl}^{(2)} &= \Pi_l^* A_{lr}^*\Pi_l A_{kr}-\Pi_l^* A_{lr}^*\Pi_k A_{lr}-\Pi_s^* A_{ls}^*\Pi_l A_{ik}\\
&\quad -A_{kr}^*\Pi_s\Pi_s A_{lr}+A_{kr}^*\Pi_l\Pi_s A_{rs}+A_{rs}^*\Pi_l\Pi_s A_{kr}\\
&\quad +\tfrac{1}{2}\delta_{kl}\Pi_s^* A_{sl}^*\Pi_r A_{rl}+\tfrac{1}{2}\delta_{kl}\Pi_s^* A_{rl}^*\Pi_r A_{sl}\\
&\quad -\tfrac{1}{2}\delta_{kl}A_{rs}^*\Pi_r\Pi_l A_{sl}-\tfrac{1}{2}\delta_{kl}A_{rs}^*\Pi_l\Pi_r A_{sl}, &(7\cdot2)\\
Y_{kl}^{(3)} &= -\Pi_l^* A_{lr}^*\Pi_k A_{lr}+A_{rs}^*\Pi_k\Pi_l A_{rs}+\delta_{kl}\Pi_r^* A_{sl}^*\Pi_r A_{sl}-\delta_{kl}A_{rs}^*\Pi_s\Pi_l A_{rs}, &(7\cdot3)\\
Y_{kl}^{(4)} &= \Pi_k^* C^*\Pi_l C-C^*\Pi_k\Pi_l C-\delta_{kl}\Pi_s^* C^*\Pi_s C+\delta_{kl}C^*\Pi_s\Pi_s C, &(7\cdot4)\\
Y_{kl}^{(5)} &= \Pi_s^* C^*\Pi_s A_{kl}-\Pi_l^* C^*\Pi_s A_{ks}-\Pi_s^* C^*\Pi_l A_{ks}+C^*\Pi_k\Pi_r A_{lr}\\
&\quad -\tfrac{1}{2}C^*\Pi_r\Pi_r A_{kl}+A_{rl}^*\Pi_k\Pi_r C-\tfrac{1}{2}A_{kl}^*\Pi_s\Pi_s C\\
&\quad +\delta_{kl}\Pi_s^* C^*\Pi_r A_{rs}-\tfrac{1}{2}\delta_{kl}A_{rs}^*\Pi_r\Pi_s C-\tfrac{1}{2}\delta_{kl}C^*\Pi_r\Pi_s A_{rs}, &(7\cdot5)\\
-\frac{2\kappa}{\hbar c}Y_{kl}^{(6)} &= \tfrac{1}{2}\Pi_k^* A_{lr}^*\Pi_l A_{lr}-\Pi_r^* A_{kr}^*\Pi_s A_{ls}-2\Pi_k^* A_{lr}^*\Pi_l A_{lr}\\
&\quad +2\Pi_l^* A_{ik}^*\Pi_s A_{ls}+\Pi_l^* A_{kr}^*\Pi_r A_{ls}+\tfrac{1}{2}A_{rs}^*\Pi_k\Pi_l A_{rs}\\
&\quad +A_{rs}^*\Pi_r\Pi_s A_{kl}-A_{kl}^*\Pi_r\Pi_s A_{rs}+A_{rs}^*\Pi_l\Pi_r A_{ks}\\
&\quad -3A_{rs}^*\Pi_r\Pi_l A_{ks}-\delta_{kl}\Pi_r^* A_{rl}^*\Pi_s A_{sl}+\delta_{kl}A_{rs}^*\Pi_r\Pi_l A_{sl}\\
&\quad +\Pi_l^* C^*\Pi_s A_{ks}-\tfrac{1}{2}\Pi_s^* C^*\Pi_s A_{kl}+\tfrac{1}{2}A_{kr}^*\Pi_l\Pi_r C\\
&\quad -\tfrac{1}{2}A_{kr}^*\Pi_r\Pi_l C+\tfrac{1}{2}A_{kl}^*\Pi_r\Pi_r C+\tfrac{1}{4}\delta_{kl}C^*\Pi_r\Pi_s A_{rs}\\
&\quad -\tfrac{1}{4}\delta_{kl}A_{rs}^*\Pi_r\Pi_s C-\tfrac{3}{16}\Pi_k^* C^*\Pi_l C-\tfrac{3}{16}C^*\Pi_k\Pi_l C. &(7\cdot6)
\end{aligned}$$

Examining $-e^{-1}j_4c_k\hbar$ and T_{k4} for the special case in which $\phi = 0$, and A_{ik} and C contain x_r through the factor $\exp(ic_rx_r)$, it is satisfactory to find that they are equal for any values of the constants γ in (6). In proving the equality, the equations

$$C = C^* = c_rA_{rs} = c_rA^*_{rs} = 0,$$

are utilized, which can be proved from the wave equations for this special case.

With regard to the Hamiltonian of such particles and T_{44}, no attempt will be made to compare them in the present paper, for a theory of second quantization for such particles in an electromagnetic field is still lacking. For second quantization of such particles in a force-free field, reference should be made to Fierz's paper (1939).

II. Particles of spin 3/2

Following Fierz & Pauli (1939) let the wave functions be c_α, $d_{\dot\beta}$, $a^{\alpha\beta}_\gamma$ and $b^\gamma_{\dot\alpha\dot\beta}$, of which the last two are symmetrical in α, β and $\dot\alpha, \dot\beta$ respectively The conjugate wave functions are c^*_α, $d^*_{\dot\beta}$, $a^{*\alpha\beta}_\gamma$ and $b^{*\gamma}_{\dot\alpha\dot\beta}$. Whenever they have their suffices explicitly written out, the *'s are omitted, since this will not cause confusion

The wave equations can be deduced as usual from varying a Lagrangian In the present case, the Lagrangian $L^{(3/2)}$ is

$$-\hbar c\{\kappa[a^\gamma_{\dot\alpha\dot\beta}b^{\dot\alpha\dot\beta}_\gamma + b^{\alpha\beta}_\gamma a^\gamma_{\alpha\beta}] - [a^\gamma_{\dot\alpha\dot\beta}\Pi^{\dot\beta\rho}a^\alpha_{\gamma\rho} + b^{\alpha\beta}_\gamma \Pi_{\alpha\rho}b^{\gamma\rho}_{\dot\beta}] + [a^\gamma_{\dot\alpha\dot\beta}\Pi^{\dot\beta}_\gamma d^{\dot\alpha} + b^{\alpha\beta}_{\dot\gamma}\Pi^\gamma_\beta c_\alpha + \text{c.c}]$$
$$- 3(d^\alpha \Pi_{\alpha\beta}d^{\dot\beta} + c_\alpha \Pi^{\dot\alpha\beta}c_{\dot\beta}) - 6\kappa(d^\alpha c_\alpha + d^{\dot\alpha}c_{\dot\alpha})\} \tag{8}$$

The wave equations are therefore

$$2\kappa b^{\alpha\dot\beta}_\gamma - (\Pi^{\dot\beta\rho}a^\alpha_{\gamma\rho} + \Pi^{\dot\alpha\rho}a^{\dot\beta}_{\gamma\rho}) + (\Pi^{\dot\beta}_\gamma d^\alpha + \Pi^{\dot\alpha}_\gamma d^{\dot\beta}) = 0, \tag{9·1}$$

$$2\kappa a^\gamma_{\alpha\beta} - (\Pi_{\alpha\dot\rho}b^{\gamma\dot\rho}_\beta + \Pi_{\beta\dot\rho}b^{\gamma\dot\rho}_\alpha) + (\Pi^\gamma_\beta c_\alpha + \Pi^\gamma_\alpha c_\beta) = 0, \tag{9·2}$$

$$2\kappa c_\alpha - \tfrac{1}{3}\Pi^\beta_\gamma a^\gamma_{\alpha\beta} + \Pi_{\alpha\dot\beta}d^{\dot\beta} = 0, \tag{9·3}$$

$$2\kappa d^\alpha - \tfrac{1}{3}\Pi^\gamma_{\dot\beta}b^{\alpha\dot\beta}_\gamma + \Pi^{\dot\alpha\beta}c_\beta = 0 \tag{9·4}$$

According to the way the notation 'h' has been used (see part I), the conjugate equations to the above are simply

$$h(b^{\alpha\beta}_\gamma) = h(a^\gamma_{\alpha\beta}) = h(c_{\dot\alpha}) = h(d^\alpha) = 0$$

At this juncture it may be pointed out that the numbers -3 and -6 in the Lagrangian were given by Fierz & Pauli as 3 and 6 which are wrong. Using the numbers 3 and 6 instead of -3 and -6, the coefficients in (9·3) and (9·4) are all positive, thus $c = d = 0$ for the case $\phi = 0$ cannot be deduced.

To get the current, vary the combined Lagrangian of the particles and the field to $\phi_{\mu\dot{\nu}}$ subject to $\delta^{\mu\dot{\nu}}\phi_{\mu\dot{\nu}} = 0$. In this way

$$\frac{1}{16\pi}\delta_{\rho\dot{\rho}}\delta^{\rho\dot{\rho}}\phi_{\mu\dot{\nu}} - \delta_{\mu\dot{\nu}}\Phi = \frac{1}{2c}j_{\mu\dot{\nu}}$$

is obtained, where Φ is a scalar to be determined by $\delta^{\mu\dot{\nu}}\phi_{\mu\dot{\nu}} = 0$ and $j_{\mu\dot{\nu}}$ is the spinor

$$-2c\frac{\partial L^{(3/2)}}{\partial\phi^{\mu\dot{\nu}}} = 2ce\{a^{\gamma}_{\rho\dot{\nu}}a^{\rho}_{\mu\gamma} + b^{\gamma}_{\rho\dot{\nu}}b^{\rho}_{\mu\gamma} - a^{\rho}_{\mu\dot{\nu}}d_{\rho} - a^{\rho}_{\mu\dot{\nu}}d_{\rho} + b^{\rho}_{\mu\dot{\nu}}c_{\rho} + b^{\rho}_{\mu\dot{\nu}}c_{\rho} + 3d_{\mu}d_{\dot{\nu}} + 3c_{\mu}c_{\dot{\nu}}\}. \quad (10)$$

It is easy to verify from the wave equations that $\delta^{\mu\dot{\nu}}j_{\mu\dot{\nu}} = 0$, so that $j_{\mu\dot{\nu}}$ can be taken as the current spinor. At the same time from $\delta^{\mu\dot{\nu}}j_{\mu\dot{\nu}} = 0$ the equation $\delta_{\mu\dot{\nu}}\delta^{\mu\dot{\nu}}\Phi = 0$ or $\Phi = 0$ is derived, and thus

$$-\tfrac{1}{2}\delta_{\rho\dot{\rho}}\delta^{\rho\dot{\rho}}\phi_{\mu\dot{\nu}} = -\frac{4\pi}{c}j_{\mu\dot{\nu}}, \quad (11)$$

as usual.

The equation for the spinor $t_{\dot{n}\dot{s}m\lambda}$ of T_{kl} is

$$\delta^{\lambda\dot{s}}t_{n\dot{s}m\lambda} + \frac{1}{c}(\delta_{mn}\phi_{\lambda\dot{s}} - \delta_{\lambda\dot{s}}\phi_{mn})j^{\lambda\dot{s}} = 0, \quad (12)$$

and the procedure of solving for $t_{\dot{n}\dot{s}m\lambda}$ is exactly the same as that for T_{kl} in the previous section or that in part 1. The equations playing the part of (9), (10) and (11) in part 1 consist of

(i) equations of the type

$$\left.\begin{aligned} a^{\gamma}_{\alpha\beta}\Pi^{*}_{\lambda\dot{s}}h(b^{\phi\psi}_{\theta}) - (\Pi^{*}_{\lambda\dot{s}}b^{\phi\psi}_{\theta})\,h(a^{\gamma}_{\alpha\beta}) = 0,\\ b^{\gamma}_{\alpha\beta}\Pi_{\lambda\dot{s}}h(a^{\phi\psi}_{\theta}) - (\Pi_{\lambda\dot{s}}a^{\phi\psi}_{\theta})\,h(b^{\gamma}_{\alpha\beta}) = 0,\end{aligned}\right\} \quad (13)$$

(ii) equations of the type

$$\left.\begin{aligned} d^{\alpha}\Pi^{*}_{\lambda\dot{s}}h(b^{\phi\psi}_{\theta}) - (\Pi^{*}_{\lambda\dot{s}}b^{\phi\psi}_{\theta})\,h(d^{\alpha}) = 0,\\ -c^{\dot{\alpha}}\Pi_{\lambda\dot{s}}h(a^{\phi\psi}_{\theta}) + (\Pi_{\lambda\dot{s}}a^{\phi\psi}_{\theta})\,h(c^{\dot{\alpha}}) = 0,\end{aligned}\right\} \quad (14)$$

(iii) equations of the type

$$\left.\begin{aligned} a^{\phi\psi}_{\theta}\Pi^{*}_{\lambda\dot{s}}h(c^{\dot{\alpha}}) - (\Pi^{*}_{\lambda\dot{s}}c^{\dot{\alpha}})\,h(a^{\phi\psi}_{\theta}) = 0,\\ -b^{\phi\psi}_{\theta}\Pi_{\lambda\dot{s}}h(d^{\dot{\alpha}}) + (\Pi_{\lambda\dot{s}}d^{\alpha})\,h(b^{\phi\psi}_{\theta}) = 0,\end{aligned}\right\} \quad (15)$$

(iv) equations of the type

$$\left.\begin{aligned} d_{\dot{\theta}}\Pi^{*}_{\lambda\dot{s}}h(c_{\dot{\alpha}}) - (\Pi^{*}_{\lambda\dot{s}}c_{\alpha})\,h(d_{\dot{\theta}}) = 0,\\ c_{\dot{\theta}}\Pi_{\lambda\dot{s}}h(d_{\dot{\alpha}}) - (\Pi_{\lambda\dot{s}}d_{\dot{\alpha}})\,h(c_{\dot{\theta}}) = 0,\end{aligned}\right\} \quad (16)$$

and their conjugates. By putting one dotted and one undotted suffix to be $\dot{n}$ and m, raising or lowering some of the suffices and performing some contractions, each of

their left-hand sides can be made the $\dot{n}m$ component of a spinor. In this way, (13) and their conjugates provide sixteen independent equations,† (14) and their conjugates eight equations, (15) and the conjugates eight equations, and (16) and their conjugates eight equations. Needless to say, there are equations obtained from the above by replacing Π, Π^* by $-i\delta$, $i\delta$, and also equations with all terms containing $\delta_{\mu\beta}\phi^{\mu\beta}$ as a factor. Multiplying them by unknown quantities λ_1, λ_2, ..., adding them to the equation obtained by substituting in (12) an assumed form for $t_{\dot{n}\dot{s}m\lambda}$, setting the coefficients of independent terms of the resulting equation to be zero and solving for the λ's and the constants in the assumed form of t, the required result is obtained.

Putting $t_{\dot{n}\dot{s}m\lambda}$ as

$$y_{\dot{n}\dot{s}m\lambda}+y_{\dot{s}\dot{n}\lambda m}+y^*_{\dot{n}\dot{s}m\lambda}+y^*_{\dot{s}\dot{n}\lambda m}, \tag{17}$$

the result is

$$y_{\dot{n}\dot{s}m\lambda}=\sum_{i=1}^{24}\gamma^{(i)}y^{(i)}_{\dot{n}\dot{s}m\lambda}+y^{(25)}_{\dot{n}\dot{s}m\lambda}, \tag{18}$$

where $\gamma^{(1)}$, $\gamma^{(2)}$, ... are arbitrary real constants and

$$y^{(1)}= a^\alpha_{\lambda\dot{s}}\Pi^*_{\alpha\beta}a^\beta_{m\dot{n}}-a^\alpha_{m\dot{s}}\Pi^*_{\alpha\beta}a^\beta_{\lambda\dot{n}}+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}a^{\alpha\rho}_{\theta}\Pi^*_{\alpha\beta}a^{\beta\theta}_{\rho}, \tag{19·1}$$

$$y^{(2)}=-\epsilon_{m\lambda}a^{\alpha\rho}_{\dot{s}}\Pi^*_{\alpha\beta}a^\beta_{\dot{n}\rho}+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}a^{\alpha\rho}_{\theta}\Pi^*_{\alpha\beta}a^{\beta\theta}_{\rho}, \tag{19·2}$$

$$y^{(3)}=-\epsilon_{\dot{n}\dot{s}}a^\alpha_{m\theta}\Pi^*_{\alpha\beta}a^{\beta\theta}_{\lambda}+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}a^{\alpha\rho}_{\theta}\Pi^*_{\alpha\beta}a^{\beta\theta}_{\rho}, \tag{19·3}$$

$$y^{(4)}= d_{\dot{n}}\Pi^*_{m\dot{s}}d_\lambda-d_{\dot{n}}\Pi^*_{\lambda\dot{s}}d_m+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}d^\beta\Pi^*_{\alpha\beta}d^\alpha, \tag{19·4}$$

$$y^{(5)}=-\epsilon_{m\lambda}d_{\dot{s}}\Pi^{*\alpha}_{\dot{n}}d_\alpha+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}d^\beta\Pi^*_{\alpha\beta}d^\alpha, \tag{19·5}$$

$$y^{(6)}=-\epsilon_{\dot{n}\dot{s}}d_\beta\Pi^{*\beta}_{m}d_\lambda+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}d^\beta\Pi^*_{\alpha\beta}d^\alpha, \tag{19·6}$$

$$y^{(7)}= d_{\dot{n}}\Pi^*_{\beta\lambda}a^\beta_{\dot{s}m}-d_{\dot{n}}\Pi^*_{m\beta}a^\beta_{\dot{s}\lambda}+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}d_\theta\Pi^*_{\rho\beta}a^{\rho\theta\beta}, \tag{19·7}$$

$$y^{(8)}= \epsilon_{m\lambda}d_{\dot{n}}\Pi^*_{\alpha\beta}a^{\alpha\beta}_{\dot{s}}+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}d_\theta\Pi^*_{\rho\beta}a^{\rho\theta\beta}, \tag{19·8}$$

$$y^{(9)}= \epsilon_{\dot{n}\dot{s}}d_\theta\Pi^*_{m\beta}a^{\theta\beta}_{\lambda}+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}d_\theta\Pi^*_{\rho\beta}a^{\rho\theta\beta}, \tag{19·9}$$

$$y^{(10)}= a^\alpha_{\lambda\dot{n}}\Pi^*_{\alpha\dot{s}}d_m-a^\alpha_{\lambda\dot{s}}\Pi^*_{\alpha\dot{n}}d_m+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}a^{\rho\alpha\beta}\Pi^*_{\alpha\beta}d_\rho, \tag{19·10}$$

$$y^{(11)}= \epsilon_{m\lambda}a^{\alpha\rho}_{\dot{s}}\Pi^*_{\dot{n}\alpha}d_\rho+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}a^{\rho\alpha\beta}\Pi^*_{\alpha\beta}d_\rho, \tag{19·11}$$

$$y^{(12)}= \epsilon_{\dot{n}\dot{s}}a^{\alpha\beta}_{\lambda}\Pi^*_{\alpha\beta}d_m+\tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{\dot{n}\dot{s}}a^{\rho\alpha\beta}\Pi^*_{\alpha\beta}d_\rho, \tag{19·12}$$

† By independent equations are meant those which do not have a linear dependence between them owing to relations of the type

$$a_\lambda b^\lambda=-a^\lambda b_\lambda,\quad a_\lambda b^\lambda c_m+a_m b_\lambda c^\lambda+a^\lambda b_m c_\lambda=0.$$

$$\begin{aligned} y^{(25)} = \hbar c\{ & -\tfrac{1}{2}a^{\beta}_{m\lambda}\Pi^{*}_{\alpha\beta}a^{\alpha}_{n\dot{s}} - \tfrac{1}{2}a^{\alpha\beta}_{\lambda}\Pi^{*}_{\alpha\beta}a_{mn\dot{s}} \\ & + \tfrac{1}{2}a_{m\lambda\dot{s}}\Pi^{*}_{\alpha\beta}a^{\alpha\beta}_{n} - \tfrac{1}{2}a^{\alpha}_{m\dot{s}}\Pi^{*}_{\alpha\beta}a^{\beta}_{n\lambda} - \tfrac{1}{2}d_{\dot{s}}\Pi^{*}_{\lambda n}d_{m} \\ & - \tfrac{1}{2}\epsilon_{m\lambda}\epsilon_{n\dot{s}}d^{\beta}\Pi^{*}_{\alpha\beta}d^{\alpha} + \tfrac{1}{2}d^{\beta}\Pi^{*}_{\lambda\beta}a_{mn\dot{s}} \\ & + \tfrac{3}{2}d_{n}\Pi^{*}_{\lambda\beta}a^{\beta}_{\dot{s}m} + \epsilon_{n\dot{s}}\epsilon_{m\lambda}d_{\dot{\beta}}\Pi^{*}_{\rho\beta}a^{\rho\dot{\beta}\beta} \\ & - 2a_{mn\lambda}\Pi^{*}_{\alpha\dot{s}}d^{\alpha} + \tfrac{1}{2}a^{\alpha}_{\lambda n}\Pi^{*}_{\alpha\dot{s}}d_{m} + \tfrac{1}{4}\epsilon_{n\dot{s}}\epsilon_{m\lambda}a^{\alpha\beta}_{\rho}\Pi^{*}_{\alpha\beta}d^{\rho}\} + \quad , \end{aligned} \tag{20}$$

$y^{(13)}, y^{(14)}, \ldots$ are $y^{(1)}, y^{(2)}, \ldots$ with a replaced by b, d by $-c$, and Π^{*} by Π, and the dots in the right-hand side of (20) denote the preceding terms treated with the same replacement.

In the case of no field, compare $-e^{-1}j_{s\dot{s}}c_{m\dot{n}}\hbar$ and $t_{n\dot{s}ms}$, where $c_{m\dot{n}}$ is the spinor appearing in the solution $\exp(-\frac{i}{2}c_{m\dot{n}}x^{m\dot{n}})$ for the unstarred wave functions For this special case

$$c_{\alpha\beta}a^{\alpha\beta}_{\rho} = c_{\alpha\beta}b^{\alpha\beta}_{\rho} = c_{\rho} = d_{\rho} = 0. \tag{21}$$

So $-e^{-1}j_{s\dot{s}}c_{m\dot{n}}\hbar$ is simply

$$-2c\hbar c_{m\dot{n}}(a^{\beta}_{s\alpha}a^{\alpha}_{\dot{s}\beta} + b^{\beta}_{s\alpha}b^{\alpha}_{\dot{s}\dot{\beta}}). \tag{22}$$

If all the γ's in (18) are zero, then

$$t_{n\dot{s}ms} = -\hbar c\{2c_{\alpha\beta}a^{\beta}_{ms}a^{\alpha}_{n\dot{s}} + c_{\alpha\beta}a^{\alpha}_{ms}a^{\beta}_{n\dot{s}} + c_{\alpha\beta}a^{\alpha}_{n\dot{s}}a^{\beta}_{ms}\}, \tag{23}$$

plus exactly the same terms with a replaced by b. Making use of (21), one has

$$c_{\alpha\beta}a^{\beta}_{ms}a^{\alpha}_{n\dot{s}} = -c_{m\beta}a^{\alpha\beta}_{s}a_{n\dot{s}\alpha} = c_{mn}a^{\alpha}_{s\beta}a^{\beta}_{\alpha\dot{s}} + c^{\beta}_{m}a^{\alpha}_{ns}a_{\beta\dot{s}\alpha},$$

$$c_{\alpha\beta}a^{\beta}_{ms}a^{\alpha}_{n\dot{s}} = -c_{\alpha n}a_{\beta ms}a^{\alpha\beta}_{\dot{s}} = c_{mn}a^{\alpha}_{s\beta}a^{\beta}_{\alpha\dot{s}} + c^{\alpha}_{n}a_{\beta\alpha s}a^{\beta}_{m\dot{s}}.$$

Similarly $\quad c_{\alpha\beta}a^{\alpha}_{ms}a^{\beta}_{n\dot{s}} = -c_{s\beta}a_{ms\alpha}a^{\alpha\beta}_{n} = \tfrac{1}{2}a^{\alpha}_{ms}(c_{\alpha\beta}a^{\beta}_{n\dot{s}} + c_{s\beta}a^{\beta}_{n\alpha}) = -\kappa a^{\alpha}_{ms}b_{\alpha n\dot{s}},$

the last equality being the result of applying the wave equation. In the same way it is found that

$$\begin{aligned} c_{\alpha\beta}a^{\alpha}_{s\dot{n}}a^{\beta}_{\dot{s}m} &= -\kappa a^{\beta}_{ms}b_{n\dot{\beta}\dot{s}}, \\ c^{\beta}_{m}a^{\alpha}_{s\dot{n}}a_{\beta\dot{s}\alpha} &= \kappa a^{\alpha}_{s\dot{n}}b_{\dot{s}\alpha m}, \\ c^{\alpha}_{n}a^{\beta}_{ms}a_{\beta\alpha\dot{s}} &= \kappa a^{\beta}_{ms}b_{n\dot{\beta}\dot{s}}. \end{aligned}$$

Hence the right-hand side of (23) reduces to

$$-\hbar c\{2c_{mn}a^{\alpha}_{s\beta}a^{\beta}_{\dot{s}\alpha} - \kappa a^{\alpha}_{ms}b_{n\dot{s}\alpha} + \kappa a^{\alpha}_{s\dot{n}}b_{\dot{s}\alpha m}\}.$$

Similarly, the right-hand side of (23) with a replaced by b reduces to

$$-\hbar c\{2c_{mn}b^{\alpha}_{s\dot{\beta}}b^{\dot{\beta}}_{\dot{s}\alpha} + \kappa b^{\alpha}_{ms}a_{n\dot{s}\alpha} - \kappa b^{\alpha}_{s\dot{n}}a_{\dot{s}\alpha m}\}.$$

Adding, it follows that

$$t_{n\dot{s}ms} = -2\hbar c\, c_{mn}(a^{\alpha}_{s\beta}a^{\beta}_{\dot{s}\alpha} + b^{\alpha}_{s\dot{\beta}}b^{\dot{\beta}}_{\dot{s}\alpha}),$$

which is exactly $-e^{-1}j_{s\dot{s}}c_{m\dot{n}}\hbar$.

The comparison of the Hamiltonian of such particles and their T_{44} is not attempted here, for a theory of second quantization for such particles in an electromagnetic field is still lacking. A theory of second quantization for such particles has been given only for the case $\phi = 0$, which can be found in Fierz's paper.

Thus the search for a symmetrical T_{kl} for elementary particles with spin ranging from 0 to 2 has been completed. The only unpleasantness in the result is that T_{kl} is not unique. Also, only the lowest possible derivatives have been considered, i.e. introduction into T_{kl} of higher derivatives of the wave functions than those contemplated here could have been made, but this has not been done.

In conclusion, the writer wishes to thank Mr K. S Kao and Mr J. T. Yang for their kind interest.

References

Fierz, M. 1939 *Helv. phys Acta*, **12**, 1.
Fierz, M. & Pauli, W. 1939 *Proc. Roy. Soc.* A, **173**, 211.
Heisenberg, W. & Pauli, W. 1929 *Z Phys.* **56**, 1
Laporte, O & Uhlenbeck, G. E. 1932 *Phys. Rev.* **39**, 187.
Tetrode, H 1928 *Z. Phys* **49**, 858.

A physico-chemical study of some induced changes in the morphology of *Bacterium lactis aerogenes*. A theory of the balance and adaptive variation of certain enzyme processes in bacteria

By C. N. Hinshelwood, F R.S. and R. M. Lodge, M A., B.Sc

(*Received* 26 *May* 1943)

[Plate 10]

A strain of *Bacterium lactis aerogenes* giving normal growth when inoculated from bouillon into a standard glucose-phosphate-ammonium sulphate medium yields many long snake-like forms when inoculated into a similar medium with a much lower glucose concentration

The size distribution in the cultures giving the snake-forms is quite different from normal, and is represented by the equation $n_l = n \exp(-l/\bar{l})$, where n_l is the number of cells of length greater than l, n is the total number and $\bar{l}$ is the mean length This holds well over most of the range but does not take into account an occasional excess, in the later stages of growth, of exceptionally long cells. This formula suggests that we are dealing with a condition where the cells elongate, but where division is delayed, and depends upon a favourable conjunction of certain independent events in the cell (If the probability of division becomes *too* small, this law itself will break down)

During the growth cycle $\bar{l}$ passes through a maximum and then decreases, most of the snake-forms disappearing again -though some occasionally persist in excess of expectation these may be forms of low viability

A size coefficient, σ, is defined which gives a good representation of the abnormality of the appearance of the culture under the microscope, and serves to characterize the distribution of lengths.

σ and $\bar{l}$ decrease as the osmotic pressure of the medium is increased by the addition of salts or of erythrite

With successive passages of a culture through the ammonium sulphate medium the power to give the snake-forms shows a regular decline and is finally lost. It does not appear to be easily restored by several passages through bouillon

The tendency to give snake-forms is enhanced by one preliminary passage through a medium containing asparagine, but growth in an asparagine medium gives normal forms.

Experiments on the effect of centrifuged medium from old cultures, of inoculum size and age, and of glucose concentration on σ and on the lag, taken in conjunction with previous work on the lag phase of *Bact lactis aerogenes*, lead to the following hypothesis: Two separate factors L and D, one of which, L, is diffusible into the medium, the other, D, being probably retained by the cells, are responsible respectively for elongation and for division D may be consumed or diluted in the process of division: and its formation may be accelerated or impeded by the presence of other substances in the medium If the cells are transferred to a new medium, the rates of formation of L and D may be out of balance and snake-forms appear. Successive passages restore the balance by a process of 'training' The mechanism of the training is discussed in the light of a hypothesis in which a crude model of enzyme synthesis having certain analogies with crystal growth is used

[*This paper has been printed in full in Proceedings B, volume* 132, *pp.* 47–67]

Pebbles, natural and artificial
Their shape under various conditions of abrasion

By Lord Rayleigh, F.R.S.

(*Received* 7 *September* 1943)

[Plates 8-10]

This paper may be regarded as the sequel to an earlier one (Rayleigh 1942) and deals with the formation of artificial pebbles under controlled conditions, and their comparison with the pebbles found in nature. As before, symmetrical pebbles having an approximate figure of revolution are chiefly considered

A series of marble pebbles, made experimentally by attrition of rectangular blocks by fragments of hard steel, is shown in comparison with natural pebbles. The series ranges from cylinders with rounded ends, to an approximately spherical figure, and then on to oblate forms ending in a disk with rounded edges These are closely matched by a series of natural flint pebbles collected from the glacial gravel Prolate or oblate spheroids, differing widely from the sphere, are not obtained in experiments of this kind, nor are they found in the gravel formation

An alternative way of making pebbles in the laboratory is by 'pothole' action. The experimental pothole is a cylindrical vessel containing water with a paddle revolving coaxially with it. The paddle maintains a vortex, which carries the stone round Either the bottom or the wall of the vessel may be made of abrasive material. When the bottom is abrasive, the pebbles are of such a shape that they lie inside a spheroid of the same polar and equatorial diameter in contrast to the previous case, where they lie outside When the sides are abrasive, the form tends to the spherical This effect appears to depend on the well-known tendency of an elongated body to set itself athwart the stream, thus the end tends to rub against the abrasive wall Spherical pebbles of considerable perfection can be made in this way.

As regards concave pebbles, the discussion in the former paper is withdrawn A new method of experiment is used, depending on the pitting of a small square of sheet glass, revolved in a box with the abrasive. If many broken flints are used, each comparable in size with the glass, the latter is chiefly pitted in the middle, tending to form a concavity. If a single flint only is used, the pittings on the polished surface are uniformly distributed, and there is no tendency to form a concavity It would seem, therefore, that concavity is produced by the edges being more protected than the middle, the protection being given by pebbles other than the one which is making the wound at the given moment. This action is apart from the rounding of the edges, which ultimately spreads, invading the concavity and producing a general convexity as in ordinary pebbles.

The present paper may be regarded as the sequel to a former one (Rayleigh 1942) and deals with the formation of artificial pebbles under controlled conditions, and their comparison with the pebbles found in nature. As before, symmetrical pebbles having an approximate figure of revolution are chiefly considered.

The vast majority of pebbles found in nature are obviously shaped by the abrasion of other like pebbles rolled with them in watercourses and on sea beaches. In the earlier work chalk was used as the material of artificial pebbles, and it was shaped by being placed in a metal box with various kinds of abrasive 'particles',* small short 'tintacks', and steel hexagon nuts. The box was kept in slow rotation about one of its shorter axes.

* For convenience this word is used even when the pieces are as large as the specimen to be abraded.

Chalk was used because it is soft enough to be readily shaped into spheroids. In the present work this material was abandoned. Marble was used instead, and it was sawn and filed into prisms of square section.

To imitate the natural process chiefly at work in the formation of pebbles, the marble block was placed in a metal can measuring $17 \times 10 \times 7$ cm. With it were placed about 250 g. of angular fragments of hard steel, weighing some 6 g. each. They were made by breaking up old files. On account of their weight and hardness, these acted as a more efficient abrasive than other pieces of marble would do, but otherwise their action is believed to be similar.

The box was rotated about its shortest axis by a small motor with a worm gear to reduce the speed. The rate was 40 r.p.m., causing the contents to fall to the opposite end of the can 80 times per minute. The power consumed was less than 20 W, so that prolonged runs could be made without too great a demand on the electric supply (from accumulators). Three or four days were sometimes necessary.

A series of pebbles was made from blocks of marble ranging from $4 \times 2 \times 2$ cm., producing a prolate pebble, through $2 \times 2 \times 2$ cm., producing an approximate sphere, to $4 \times 4 \times 2$ cm., producing an oblate pebble. The finished pebbles were of course considerably smaller.

The actual pebbles are shown in plate 8, figures 1–8. They were attached to a glass plate and printed by a simple shadow method with a distant arc lamp as source, and without using lenses. They are reproduced in actual size.

As regards the size of the initial rectangular prisms, these were.

(1) $4 \times 2 \times 2$ cm.,	(2) $3{\cdot}3 \times 2 \times 2$ cm.,	(3) $3 \times 2 \times 2$ cm.,
(4) $2{\cdot}7 \times 2 \times 2$ cm.,	(5) $2 \times 2 \times 2$ cm.,	(6) $0{\cdot}2 \times 2{\cdot}5 \times 2{\cdot}5$ cm.,
(7) $2 \times 3 \times 3$ cm.,	(8) $2 \times 4 \times 4$ cm.	

Below the row of artificial pebbles are placed a series of natural pebbles of flint (plate 8, figures 9–15). These also have an approximate figure of revolution; the axis is shown vertical when the printed matter on the diagram is oriented for reading.

Flint pebbles of the desired symmetry are not very common—indeed, not more than 1 or 2 % of the local flints have any regularity at all, and the percentage having symmetry about an axis is far smaller still. It has not therefore been possible to match the artificial series of pebbles exactly with natural ones. Nevertheless, it will probably be agreed that the general resemblance of the two series proves that the artificial process is representative of the formation of natural pebbles.

There is no direct knowledge of the initial shape of the angular piece from which a particular rounded pebble was formed in nature. The pebbles here considered have an axis and a plane of symmetry, and they would most naturally be regarded as having originated in an approximate prism, more likely a square prism than any other, for triangular or pentagonal prisms, for example, are unlikely to occur. An initial solid having less symmetry than these can hardly satisfy the require-

ments, at least if the abrasion is by friction with other pebbles. As will be seen, the case is different if abrasion in a pothole is taken into account.

It will be seen from the extremes of the series 1–8 (plate 8) that it is quite illusory to regard the ellipsoid as a typical form for a pebble produced by abrasion by other bodies of comparable size or smaller size than its own. The axial section of such a pebble always lies outside the ellipse, or, in other words, the pebble allows a spheroid to be inscribed within it, touching only at the poles and at the equator.

The nearly spherical form in the middle of the series is of course an exception to this, and it is obvious *a priori* on general geometrical grounds that a pebble deviating only slightly from the sphere can be closely represented as a spheroid. Moving further along the series it will be noticed that this approximation breaks down and can have no physical significance. At the ends of the series there is no kind of resemblance to a spheroid. No 1 may be described as a cylinder with rounded ends, no. 8 as a disk with rounded edges.

In the cases just treated, simple inspection shows that those pebbles with figures of revolution which deviate considerably from the sphere are not spheroidal, but, if fitted to a spheroid at the poles and the equator, will bulge away from it at intermediate latitudes.

To examine this matter more closely, the actual size silhouette was enlarged photographically on to glass of half-plate size. The glass enlargement was put down as symmetrically as possible, film side downwards, on to a piece of squared paper with mm. rulings, on which the proposed axes had been marked. The glass was then secured to the section paper with sealing wax to prevent relative movement The ordinate corresponding to a given abscissa was read off in all four quadrants and the arithmetic mean was taken. In this way, a single quadrant of an idealized pebble section was defined and plotted on the arbitrary scale of the enlargement. The corresponding quadrant of an ellipse with the same axes was calculated and plotted for comparison on the diagram. The scale of the diagrams is arbitrary, and not the same for all.

The ellipse is shown by a full line, and the pebble outline by a dotted line, when it differs widely enough from the ellipse to admit of separate representation.

Figures 16 and 17 show artificial marble pebbles, prolate and oblate, from the above series. In both cases the principal section of the pebble lies outside the ellipse drawn to the same axes, and I have not so far found any exception to this rule among artificial pebbles shaped by mutual attrition, or among natural pebbles (figures 18, 19). I have, however, met with a pebble with shape approaching much more closely to the oblate spheroid. This pebble was one of a collection given me by Sir D'Arcy Thompson, and was figured in the previous paper (Rayleigh 1942). It was perhaps from the river Spey, but this is uncertain.* It is very unfortunate that no precise details are available as to the origin of these pebbles.

* Other oblate pebbles in the same collection, but not of the same material, show more nearly the usual flattened form.

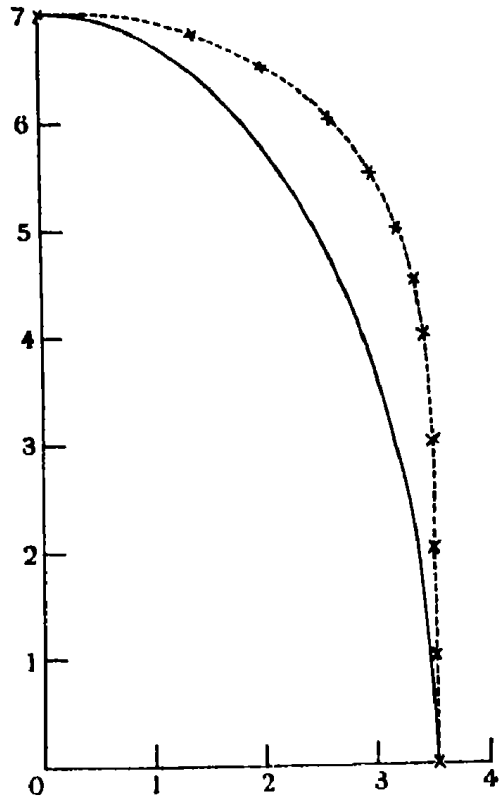


FIGURE 16

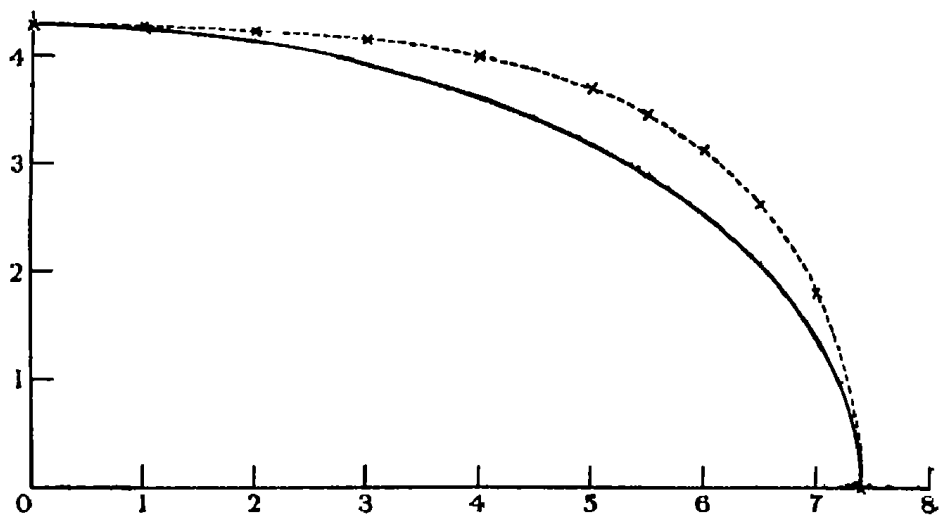


FIGURE 17

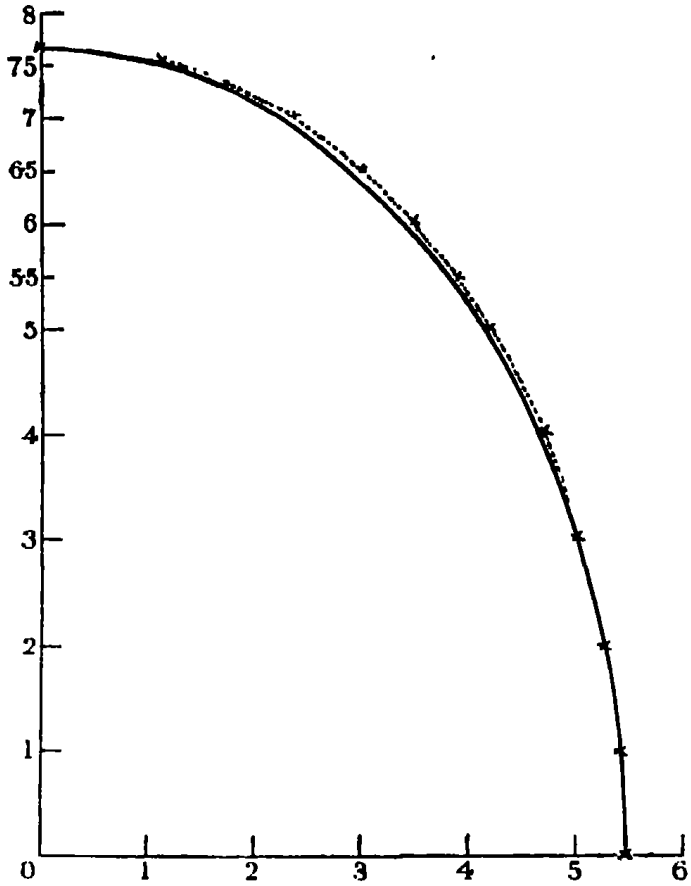


FIGURE 18

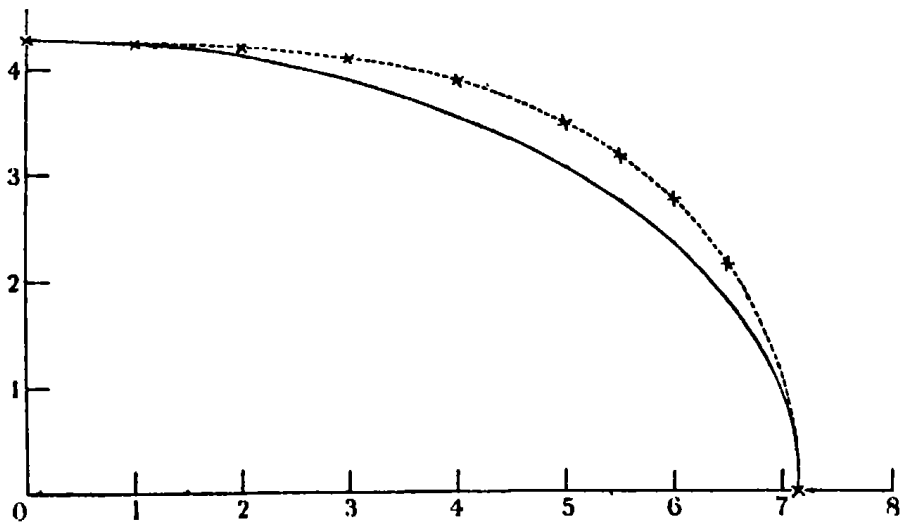


FIGURE 19

Reflecting, however, on the problem of how such an ellipsoidal pebble might be made in nature, abrasion on rocky river beds or in potholes was considered.

It is evident that the abrasion of a pebble in these occurs under conditions very different to those used in the earlier experiments, in which the pebble was abraded by bodies of the same dimension as its own or less. When it is abraded on a river bed or in a pothole, it is, in effect, being abraded by a fixed body of dimensions very large compared with its own. The idea was conceived of making an artificial

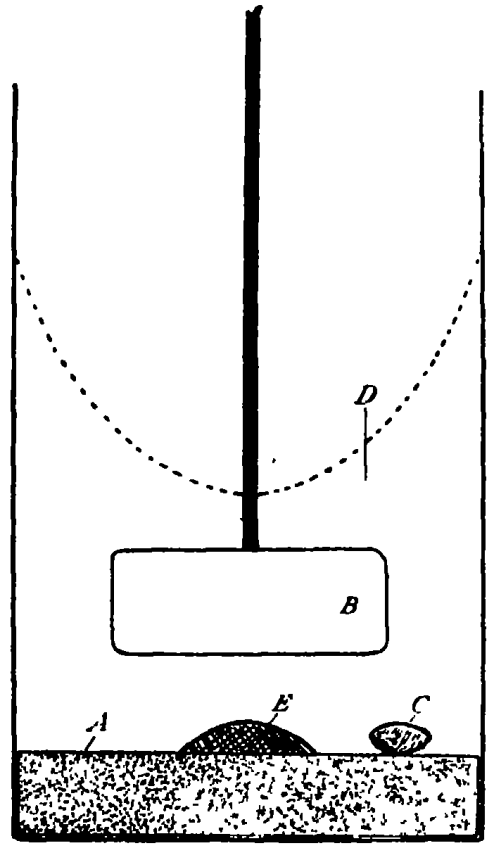


FIGURE 26

pothole in the laboratory. I began with an ordinary flower pot with water entering tangentially from a hose connected to the water supply, and flowing out through the central hole. This approximates to what Rankine (1864, p 574) calls a free vortex. Small pebbles of chalk were carried round and abraded by the unglazed earthenware surface. The angular velocity is greatest at the centre, and a short perforated zinc tube was placed axial with the pot to prevent the pebble getting into the exit hole and plugging it This arrangement gave suggestive preliminary results, but was wasteful of water and not very effective. A forced vortex, i.e. one in which the water is caused to rotate with nearly uniform angular velocity, was then adopted. The vessel, a cylinder 27 cm. in diameter (figure 26), was partly full of water which was kept in rotation by a revolving paddle *B*. To abrade the

experimental pebble a stationary abrasive disk A covered the bottom of the cylinder. This disk consisted of carborundum grit, i.e. crystals of about 1 mm. linear dimensions, mixed with about one-third its bulk of Portland cement, and allowed to harden for several days. This abrasive cement was worn away to a considerable extent in the course of the work. It answered the purpose sufficiently well, but a coarse vitrified carborundum wheel as used in grinding machines would probably have made a better (stationary) disk.

It was necessary to prevent the stone getting to the centre of the vortex, where it would be stationary on the bottom. A hemispherical lead weight E of 9 cm. diameter was placed in the middle of the disk to prevent this, and the stone C was limited to move in the annular track between this and the side of the cylinder. As soon as the paddle was started it went round this course, turning on to every side, edge and corner in succession as it was dragged over the abrasive surface by the water current.

It constantly knocked against the side of the cylinder, and it was possible to judge from the sound when all was working satisfactorily, even when the water had become opaque with the mud resulting from the grinding.

The size of pebble that can be moved will depend—other things being equal—on the circumferential velocity v of the water. If the water revolves with uniform angular velocity, the pressure will be greatest at the outside, the water surface being a paraboloid. The height of the surface above the central depression is $v^2/2g$, when g is the acceleration of gravity. For a velocity of 2 m./sec. the height is 20·4 cm. The actual mean speed was nowhere quite so large.

To move a rectangular block and to overturn in from its stable position is of course more difficult than to move a rounded pebble which can roll. My arrangement would keep a marble block $6 \times 6 \times 3$ cm. weighing 320 g. on the move It was found, however, that pebbles from a block as large as this did not preserve a very satisfactory degree of symmetry. The size ultimately preferred was about $4 \times 4 \times 2$ cm. to produce oblate pebbles, and $4 \times 2 \times 2$ cm. to produce prolate ones. These sizes had been used for the pebbles produced by other methods.

It is possible that by planning everything on a larger scale and with higher angular velocities, perfect pebbles of large size might be produced in the vortex on an abrasive bed.

Coming now to the results, the plate shows that the pebbles formed by this method (figures 22, 25) are of quite a different shape from those produced by attrition with steel fragments (figures 21, 24). Their meridional sections were plotted in comparison with an ellipse in the way already explained, and it will be seen that the pebble, whether prolate or oblate, lies inside the spheroid which is adjusted to fit it at the poles and at the equator (figures 27, 28).

It will be seen then that the most common process, attrition by other pebbles, produces a pebble which lies outside the spheroid. Another process, of the general nature of attrition in a pothole, produces pebbles which, though nearer to the spheroid than the common kind, lie inside it instead of outside. Such pebbles can

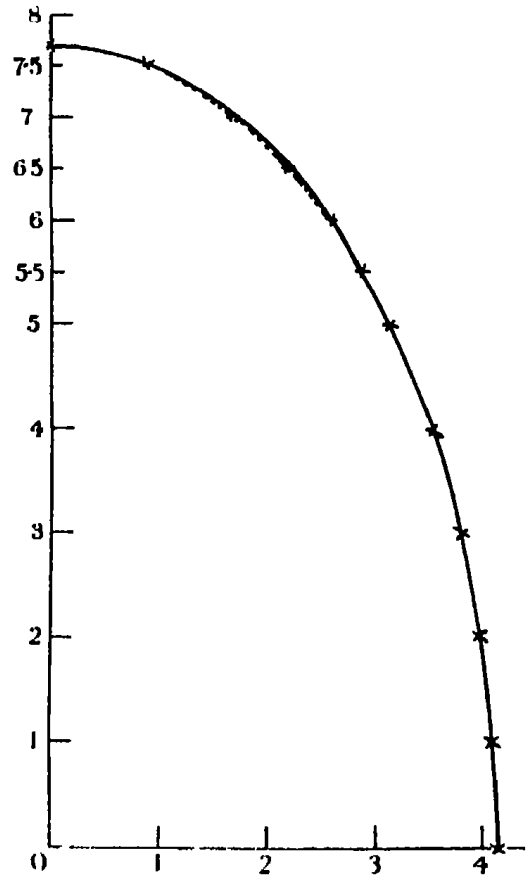


FIGURE 27

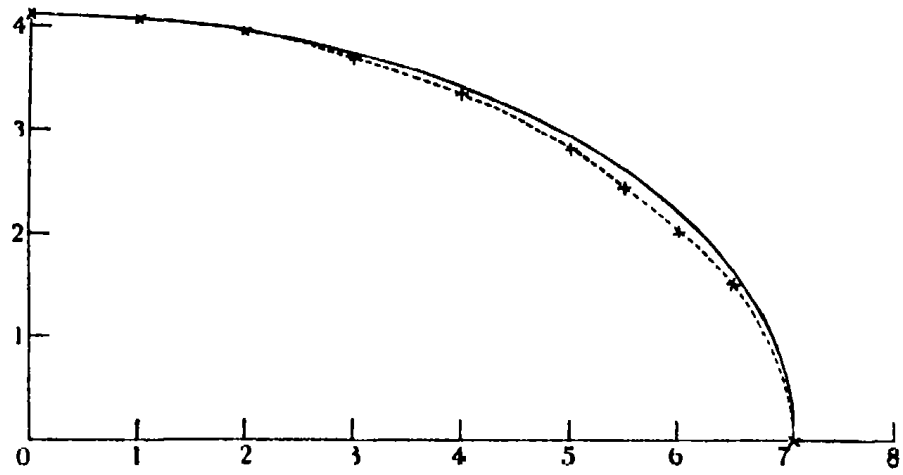


FIGURE 28

readily be produced in experiment. I have not so far met with any natural ones, but better opportunities of search may lead to the discovery of such.

Clearly the two processes may have co-operated in any given case, and the occasional occurrence in mountainous districts of pebbles showing a close approximation to the spheroid can be explained, like the one figured in the previous paper (Rayleigh 1942, facing p. 116).

The artificial pebbles produced by the pothole method, as it may be called, are in general more perfectly shaped than those produced by attrition with steel fragments. Even unsymmetrical natural pebbles of flint or quartzite tend rapidly to acquire symmetry if they are worked down in the vortex. Attrition by other pebbles does much less to correct asymmetry

The only actual potholes which I have had the opportunity to examine are some in the sandstone of the North Tyne The pebbles therein were, however, too irregular in shape to be of any use for the present purpose.

Spherical pebbles

These present an intriguing problem, on which, however, some additional light has been obtained. It was mentioned before (Rayleigh 1942, p. 107) that spherical pebbles are occasionally met with in the neighbourhood of Terling, Essex, in the glacial gravel. About six of them have come to my notice in the course of more than a year's interest in the subject. The surface is not very smooth. The size is in general rather large, 4–8 cm. in diameter, not favouring the view that this spherical shape is due to unusually prolonged attrition. The approximation to a sphere is such that the ratio of maximum to minimum diameter ranges from about 1·05 to 1·08, disregarding small local depressions These seem to stand apart from ordinary quasi-spherical pebbles, for which the corresponding ratio is not often less than 1·2. The indication seems to be that these spherical pebbles are a class apart, not to be regarded as resulting from the modification of those which are less spherical. I shall return to the experimental aspect of this presently.

On the observational side, Mr J. Reid Moir, F.R.S., kindly drew my attention to a paper by F. N. Haward (Haward 1924), 'On the origin of the battered spherical flints called cannon shot from the glacial gravel of Norfolk'. Haward takes the view that these flints were not shaped spherical by abrasion, but that they were of spherical form *in situ* in the chalk when they originated, and he figures one of $2\frac{3}{4}$ in. diameter from the chalk pit, Herringswell. Mr Reid Moir kindly obtained a specimen of these 'cannon shot' from the Norfolk gravel, and on comparing its general appearance with those found hereabouts, I have no doubt that they are of the same origin. It appears therefore that these spheres were not formed by abrasion at all, and there is no further difficulty on this head.

There remains the problem of the spherical quartzite pebbles mentioned before (Rayleigh 1942, p. 108) from Woolshed, Beechworth, Victoria, Australia (Dunn 1911, plate 7). The one or two specimens in the Natural History Museum figured

by Dunn and presented by him are strikingly accurate in shape. I have not so far been able to get any further information about these pebbles or the conditions in which they are found. It would be important to know how far the spherical ones are common or exceptional in the locality. I have not been able to hear of them in other museums. Lady Rayleigh informs me that long ago she found a perfect sphere some 2 in. diameter of some brown stone in a pothole in Inverness-shire. No other similar stones were found at the same time. The locality was Invergloy on the river Spean, Loch Lochy.

The problem was to determine how good an approximation to a sphere could be produced starting with a stone moderately oblate or prolate. The experiment already described (Rayleigh 1942) did not favour the idea that abrasion with other stones could produce a good sphere, but special tests were made starting with a piece of marble nearer the sphere than before, and abrading by hard steel fragments.

shape	state	length cm.	diameter cm.	length/ diameter
oblate spheroid	initial	1 88	2·35	0·840
	after abrasion	1·30	1·66	0·782
prolate spheroid	initial	2·13	1 90	1·11
	after abrasion	1·53	1 36	1·13

The oblate spheroid becomes more oblate. The prolate spheroid alters very little, but if anything becomes more prolate. Thus there is no tendency towards the sphere.

The next tests were made by working down spheroids slightly oblate, and slightly prolate, in a vortex on a flat abrasive bed, as described above:

shape	state	length cm.	diameter cm.	length − diameter
oblate spheroid	initial	1·95	2 09	0·934
	after abrasion	1 02	1·14	0·895
prolate	initial	2·13	1·98	1·075
	after abrasion	1·07	1·00	1·07

Again the ratio length/diameter remains sensibly unaltered, even when a large reduction of diameter takes place. Thus there is no progress towards the sphere when abrasion is in the vortex on a flat bed.

It seemed clear therefore that the smooth spherical pebbles in Dunn's collection could not have originated in either of these ways, and I was somewhat at a loss. The artificial pothole so far employed has an abrasion *bottom* and smooth *walls*. The natural pothole has both bottom and walls abrasive, the bottom being curved like a more or less hemispherical bowl. At the suggestion of Dr E. B. Bailey I tried the use of vertical abrasive walls. A cylindrical ring or hoop of carborundum concrete was made 13 cm. inside diameter and 3 cm. length in the axial direction. This was placed on the bottom of a metal drum and covered with water to a depth of 6 cm. or more. A central paddle was rotated by an electric motor as before. A

roughly shaped pebble of marble lay on the bottom, and when the paddle was going it rolled round the periphery against the side walls (figure 29). The pebble was measured from time to time with vernier calipers, and there was no difficulty in determining the maximum and minimum diameter to an accuracy of $\pm 0{\cdot}1$ mm.

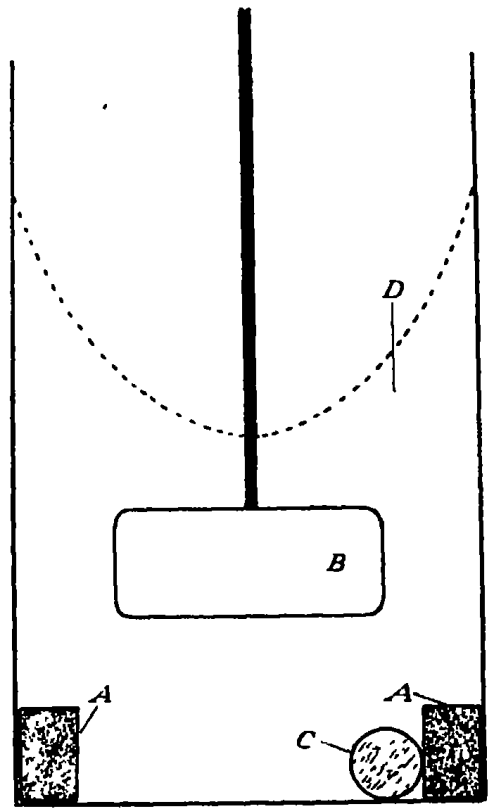


FIGURE 29

The ratio of these diameters was taken and its excess over unity measures the departure of the pebble from a sphere:

weight g.	maximum diameter cm.	minimum diameter cm.	differences cm.	difference of diameter ÷ min. diameter
19·7	2·52	2·31	0·21	0 083
17·2	2·39	2·24	0·15	0·067
15·6	2·29	2·17	0·12	0 055
13·8	2·19	2 09	0·10	0·048
12·7	2·13	2·05	0·08	0 039
11·8	2·06	2·00	0·06	0 030
9 1	1·88	1·82	0·06	0 030

The above experiment took about three days run. The original rather irregular lump has been reduced in this time to an approximate sphere very satisfying to the eye, its error of figure (3 %) being scarcely perceptible by inspection, though of

course conspicuous enough when calipers are used. It appears that the limit of accuracy has been reached, and that a longer run would not have improved it.

No ordinary pebble found in watercourses or in gravel deposits is nearly as good as this one, which has the aspect of a manufactured article. The best pebble in the Dunn collection above mentioned is about as good. A larger sphere, 3 cm. in diameter, and of nearly the same accuracy, was produced in the laboratory in the same way (plate 9, figure 30).

Now as to the explanation of this effect. It is known that an elongated body pivoted in the middle sets itself across a liquid stream. This is readily verified and may be taken as an experimental fact. From the point of view of theoretical hydrodynamics the problem is difficult, and so far as I know the only case where it has been solved is that of two-dimensional motion past an oblique plane lamina. It was shown by the work of Kirchhoff and the late Lord Rayleigh (Rayleigh 1876) that the centre of pressure is in such a position that the resultant pressure has a moment tending to set the lamina across the stream, and this example is a sufficient illustration of the general principle which applies in cases less mathematically tractable. The idea suggested is that the pebble, though carried round by the vortex, may be regarded as to some extent pivoted at the point at which it rests momentarily on the bottom of the vessel. It will therefore tend to set itself radially while moving circumferentially on the bottom. Further, centrifugal force will cause the *end* of the elongated pebble to press against the abrasive walls, and as it does so it will be worn away. It appears then that the longest axis places itself in the position which is favourable for shortening it, and this process of continually grinding away the longest dimension must clearly tend to produce the spherical form. It is perhaps scarcely to be foreseen how far this process would go in producing a perfect sphere. It was found, as above, that the process ceases to be effective under the conditions used when the longest axis is 3 % in excess of the shortest. It will probably cause surprise that a body so little elongated as this should have any appreciable tendency to set itself across the stream. Nevertheless, a specially designed experiment, though only qualitative, shows clearly enough that such a tendency *is* in evidence. An approximate spheroid of boxwood was turned in the lathe, 2·06 cm. long and 2 cm. in diameter. It was carried on the middle of a brass torsion wire its length being perpendicular to the wire. The control was reasonably stiff, giving a period of complete vibration of about half a second. The motion was of course damped by the water when the ball was immersed. The long dimension was marked in black and set itself radially when the boxwood ellipsoid was held in the vortex, with the torsion wire vertical, even when the velocity was less than that used in shaping the stones. It appears therefore that the directive action of the stream remains appreciable even when the ball is approaching the spherical form.

When the ellipsoid is more elongated, say to the extent of 5 % instead of 3 %, the directive force is stronger, and can show itself when the ball is mounted on a vertical pin working in a socket with considerable solid friction, instead of the

torsion wire. The transverse radial position of the ellipsoid is then held by the solid friction, and can be verified further (if judged necessary) when the motion of the water has been stopped.

In order further to check the correctness of this view of the matter, a small square prism of marble 9 mm. square and 15·7 mm. long was placed in the vessel inside the abrasive cylinder. The prism had developed rounded ends, and in 4½ days the all-over length was reduced to 12·7 mm., the transverse measurement being almost unaffected. A few days more would doubtless reduce it to a sphere.

This experiment was carried out on a small scale to reduce the linear thickness to be abraded away, and hence the time required. With very long and narrow prisms this experiment does not succeed well. The reason for this is not difficult to guess, but seems hardly worth going into at length.

Pebbles with concave surfaces

I gave an account of these in the former paper. Tabular flint pebbles often have a slight concavity, and it was shown that this could be imitated experimentally by abrading a flat surface of steatite with fragments of metal of a size comparable with its own, but not with relatively small particles such as gunshot.

In further experiments, I have used marble sheets or slabs of *c.* 10 mm. thick as found in commerce. A piece of 3 cm. square and 8·80 mm thick was taken and gauged at the centre of each side and in the middle The mean thickness was 8·90 mm It was then placed in a can with sharp broken flints of dimensions comparable with its own and covered with water

After prolonged rotation (some 24 hr.) the piece was reduced to a cushion shape with hollow faces, easily recognizable as such by the use of a straight edge, or perhaps even without it The concavity (exaggerated) is indicated in figure 31

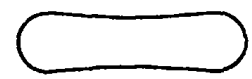


Figure 31

The thickness (at the edges) when it is a maximum averages 7·90 mm. and in the middle 7·78 mm. The difference, 0·12 mm , seems to have reached a limit, not increasing toward the end of the run It is only a small fraction of the total reduction of thickness, which is about 1·0 mm.

I discussed the cause of this concavity in the previous paper, but the view then taken was quite inconclusive, as pointed out to me by Dr E. B. Bailey, and I wish to withdraw it. In attacking the subject again, it was resolved to check every step as far as possible by experiment. In determining the effect of abrasion by impacts on a plane surface, the idea suggested itself of using a polished surface of glass with a substance harder than glass, such as flint or carborundum, for the impinging body. Each effective impact should then make a highly localized pit in the polished surface. The distribution of these pits can be examined under a reading glass or other low-power magnifying lens. This method supplements observation of curvature. It gives a much quicker result, and exhibits the action in a fresh

light. The conditions of abrasion were as described above. A new piece of plate glass of a standard size, 3×3 cm. with a thickness of about 3 mm., was used in each test.

Such a piece was abraded for a few minutes with about 60 g. of carborundum grit, grain size about 1 mm. The result of this experiment was that the pittings of the glass were concentrated at the edge, and were fewest at the centre (plate 10, figure 32). This is the common process of producing convexity in pebbles, the main and conspicuous result of abrasion which is commonly noticed.

In the next experiment, some twenty broken flints, each one of dimensions comparable with that of the glass, i.e. 3 cm., were put in instead of carborundum. It was now found that the pittings were concentrated at the centre (plate 10, figure 33). At the same time and quite distinct from this, some chips of relatively larger size were broken off from the edges of the glass. It is evident that the pittings concentrated at the centre represent the beginnings of a concave curvature there. The large chips at the edge represent the beginning of a rounding of the edges, which, if the process were continued long enough, would spread towards the centre, overwhelming and obliterating the concave curvature, and eventually replacing it by convexity. The large chips detached while the edge of the glass is sharp would be replaced by pittings when it ceased to be so; there would then be a maximum of pittings at the edge and at the middle, with an intervening minimum, which latter represents the high ridge near the periphery of a specimen such as that represented diagrammatically in figure 31. As abrasion proceeds, the edges get worn away, and the edge abrasion continually encroaches on the concavity.

It is hoped that this point has been clearly made out. It remains to consider why concavity is produced at all. I may confess that it was to me an unexpected result of abrasion, and I have not been able to find that it has been noticed by previous writers.

As is seen in this and the previous paper, concavity is only noticeable when the abrasive is in pieces comparable with the size of the specimen, and is not produced by a powdered abrasive. To simplify the conditions, only one single piece of flint was used as abrasive, comparable in size with the glass. In this case there was definitely no tendency for the pittings to be concentrated in the middle (plate 10, figure 34). It is concluded therefore that this tendency is due to the action of pieces other than the one which actually makes the effective blow, and this action can hardly be anything but a *protective* one. In other words, when there are a number of abrasive pieces, these tend to cover up the edge of the specimen more often than the middle, leaving the latter more likely to receive a wound. Hence the concentration of the pittings in the middle. The tendency is not a very marked one, and consequently the concavity never becomes deep, the countervailing tendency to produce a convexity from the edge inwards being the more powerful, and eventually asserting itself everywhere. The concavity is thus not a very important or significant effect, but, since it presents itself in nature, even without the experimentalist, it seemed to be worthy of examination.

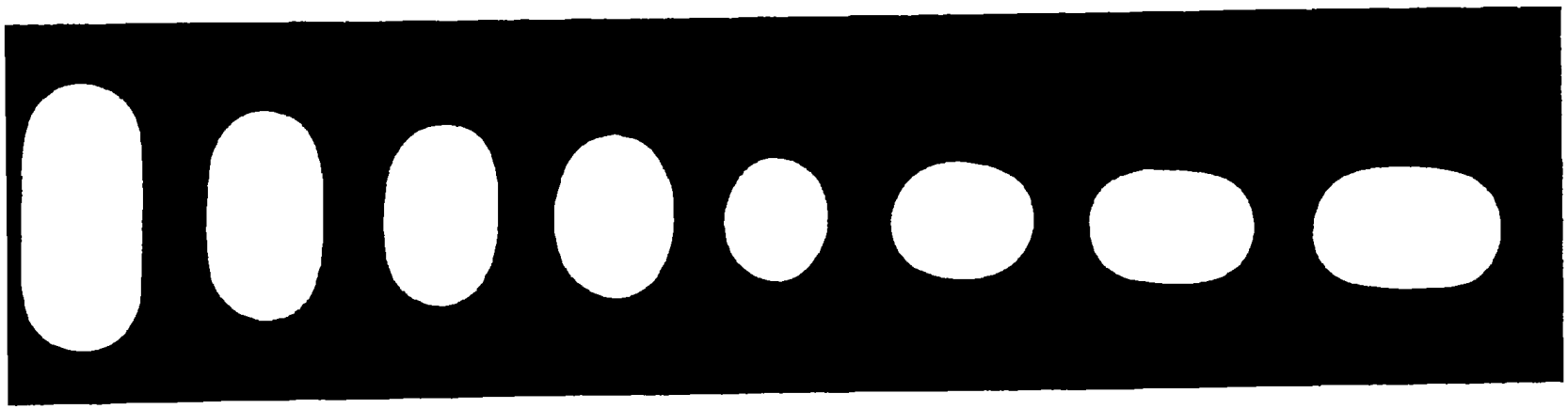
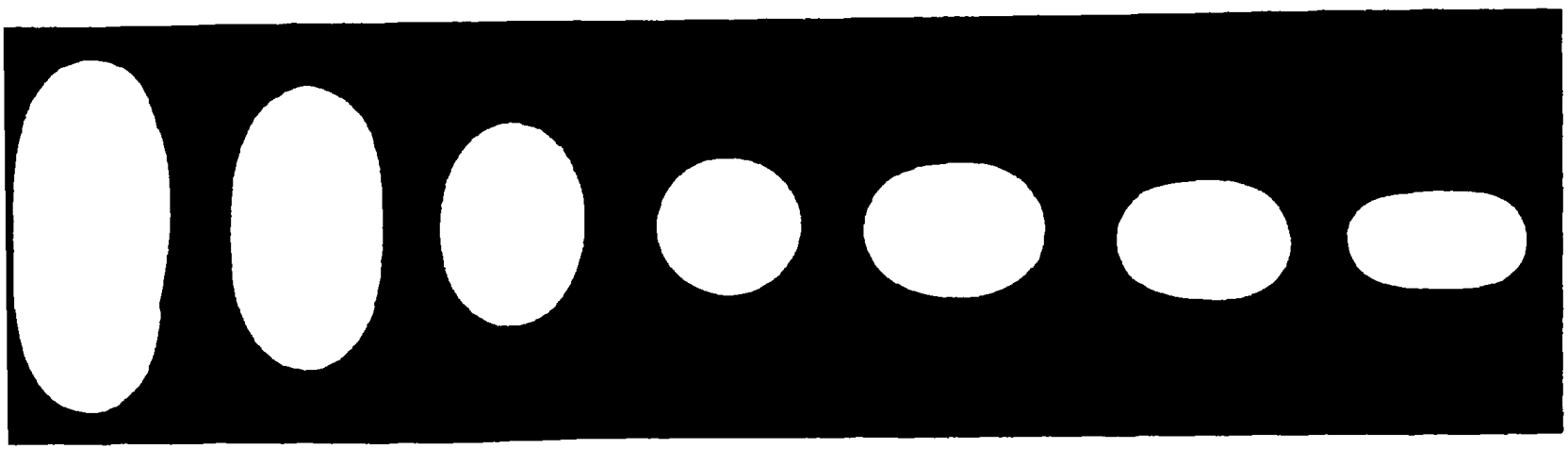


FIGURE 1–8 Artificial pebbles of marble.



FIGURES 9–15. Natural pebbles of flint

These are all approximate figures of revolution. Axis vertical thus ↑

REFERENCES

Dunn, E. J. 1911 *Pebbles*. Melbourne.
Haward, F. N. 1924 *Proc. Geol. Ass.* **35**, 122.
Rankine, W. J. M. 1864 *Applied mechanics*. London.
Rayleigh, Lord 1942 *Proc. Roy. Soc.* A, **181**, 107–118.
Rayleigh, Lord (the late) 1876 *Phil. Mag.* **2**, 430–440; *Collected papers*, **1**, 287–296.

The combination of fibrous proteins with acids

BY G. A. GILBERT AND E. K. RIDEAL, F.R.S.

Laboratory of Colloid Science, Cambridge

(*Received* 4 *October* 1943)

An approximate titration equation is developed for the reaction between fibrous proteins and acids, in which account is taken of the valency and intrinsic affinities of the acid anions, and of the limited number of adsorption sites available for anions. The transition from soluble to fibrous proteins is traced in order to demonstrate that the main difference between the titration curves of the two types of protein arises from the difference in potential developed during the adsorption of protons.

Recent work on the dissociation curves of soluble proteins has shown that their main features can be explained if it is assumed that salts such as protein chloride or sodium proteinate are completely dissociated in solution (Cannan 1942). The difference between the number of protons reversibly combined with a dissolved protein molecule, and the number of potentially negative groups such as carboxyl and phenolic hydroxyl, then determines the net charge on the molecule. If the number of combined protons is plotted against pH, a complete titration curve is found to contain a series of steps corresponding to the successive dissociation of different types of group (guanidine, amino, imidazole, carboxyl, for example), the position of the steps being determined by the net charge of the whole molecule, and the intrinsic dissociation constants of the groups.

An acceptable theoretical analysis of these curves was first given by Linderstrøm-Lang (1924), and has since been extended by other workers (Cannan 1942; Cannan, Palmer & Kibrick 1941, 1942). For details, the references just indicated should be consulted, but, to put the matter briefly, these authors assume that a protein molecule is spherical, with its charged groups distributed at random over its surface. If the molecule gains a proton its potential is raised by an amount e/Dr and the electrostatic potential energy of each of the protons already adsorbed is increased by roughly e^2/Dr, where D is the dielectric constant of water, e is the electronic charge, and r is the radius of the molecule. Shielding by the ion atmospheres surrounding

the polyvalent protein ions and the free protons is computed with the help of the theory of Debye & Hückel (1923), and the statistical factor that appears in the dissociation of polyvalent ions is calculated after the manner of Adams (1916) and Bjerrum (1923).

Soluble proteins are mentioned here to enable us to show that the transition from a soluble protein to a compact fibrous protein must be accompanied by a change in the form of the titration curve. The same principles apply over the whole of the complex curve, but they will be illustrated with respect to the part of the curve for which carboxyl groups are responsible, since the dissociation constants of the latter are approximately the same as one another, and are sufficiently different from those of basic groups for this part of the curve to be almost homogeneous. In addition, we shall consider a protein in which the carboxyl groups are equal in number to the basic groups (since this has been found to be approximately true for the protein, wool keratin, in which we are particularly interested in this paper), but once again, it is easily seen how the treatment can be modified to suit a more general case, even if the result is algebraically more complicated. It follows from these conditions that at the isoelectric point all the carboxyl groups are dissociated, and all the basic groups charged.

As acid is added to such an isoelectric dissolved protein, protons are taken up with decreasing affinity as the protein molecule assumes a positive charge proportional to the number of protons adsorbed. Therefore, instead of this part of the titration curve being identical with the neutralization curve of a mono-carboxylic acid, it is spread over a wider range of pH. The maximum potential that can be built up on the molecule is limited to that produced by the positively charged basic groups when all the carboxyl sites have been neutralized (although it is not implied by this that there is any differentiation between the different adsorbed protons as far as this potential is concerned). For many naturally occurring soluble proteins it is reasonable to suppose that the positive charge is never high enough (remembering the shielding effect of ion atmospheres) to bring about more than a very slight adsorption of unspecific ions like the chloride ion, from dilute solutions of acid. Egg albumin, for instance, is limited to a positive charge of about forty at the most (Cannan *et al.* 1941). A high concentration of anions in solution, or the presence of other sources of attraction beyond these coulombic forces, leads to a very different situation, and anion adsorption is then quite likely.

It is found that titration curves obtained at constant ionic strength are steeper, the higher the ionic strength (Cannan 1942). One school of thought (Abramson, Moyer & Gorin 1942) attributes this to the adsorption of ions other than protons in amounts increasing with ionic strength, but Cannan *et al.* (1941) have calculated the magnitude of the effect very convincingly by assuming complete anion dissociation throughout, and merely increased shielding with rising ionic strength.

We can now indicate how the difference in dispersion of fibrous and soluble proteins leads to differences in their titration curves. Instead of being limited to some forty or so, the number of protons that a single compact element of fibre can absorb

is very great. For a protein molecule of radius r in which the proton sites are distributed uniformly and at a density independent of r, the potential increment produced by each proton decreases according to $1/r$, but the number of sites increases with r^3. In fact, if r is of microscopic, instead of atomic dimensions, only a minute fraction of the proton sites can be filled before a potential is reached that makes any further selective adsorption of protons quite impossible. After this, each adsorbed proton must be accompanied by an anion, and furthermore, the anion must penetrate into the interior, and not merely form a tightly adsorbed layer at the fibre surface, for such a separation of anions and protons would still lead to prohibitive potentials.

This picture for the familiar stoichiometric adsorption of anions and cations by fibres leads one to expect a fibre in acid solution to be at a higher positive potential than a soluble protein that differs from it only in size of particle. In qualitative agreement with this, the acid-titration curve of wool keratin lies in a region of much lower pH than the corresponding curve for a protein such as egg albumin (Steinhardt & Harris 1940). It should be noticed that the apparent increase in the acid strength of the carboxyl groups in a fibre is not attributed to the presence of positive groups in close proximity to each carboxyl group, but to the positive potential of the fibre as a whole. There is therefore not a very close analogy with amino acids like glycine, where the interaction between neighbouring charged centres is high because they are separated, not by water of hydration as in salt links, but by a chain of carbon atoms with a low dielectric constant.

Most of the fall in potential between the fibre and solution probably takes place in the diffuse ion atmosphere surrounding the fibre, just as in the case of soluble proteins, although now the anions in this layer are a negligible part of the whole. The surface layer of tightly adsorbed anions contributes another part, and the rest occurs within the fibre itself, since thermal disturbance of the excess protons prevents them being entirely in the outermost sites. The final rise within the fibre, which is analogous to the fall in the external medium, will probably be sharp enough for the adsorbed ions to be treated as if they were all in an equipotential region. Measurements of electrophoretic mobility can only give information about the outer diffuse layer.

The acid-titration equation for protein fibres

If the fibre is compact, the most likely sites for the adsorbed anions are the positively charged basic groups, almost all of which in wool keratin are usually supposed to be paired with carboxyl groups (Speakman & Hirst 1933). In the model on which calculations will be made, it will be assumed that all these positive groups have identical properties as far as their interaction (intrinsic affinity) with anions is concerned, and that all the carboxyl groups are alike in their affinity for protons. Further, the important assumption will be made that an anion is free to occupy any positive site, irrespective of whether or not the positive site is next to a carboxyl

group that has been neutralized by a proton. Only one anion, however, can be adsorbed by any one site, just as the adsorption of one proton makes a carboxyl site no longer available. Anions and protons are thus regarded as being adsorbed independently, except in so far as they influence one another through their contribution to the net charge of the fibre.

The dependence of the chemical potential of an uncharged adsorbed substance, distributed at random among a limited number of sites, on the fraction θ of the sites occupied, has been deduced statistically by Fowler (Fowler & Guggenheim 1939), and is given by

$$\mu = \mu^0(TP)_{\text{fibre}} + RT \ln \frac{\theta}{1-\theta}, \tag{1}$$

where $\mu^0(TP)$ is the chemical potential of the substance at a pressure P and temperature T when $\theta = 0{\cdot}5$, and R is the gas constant. If the substance has unit charge, and the sites are in a region of potential ψ, the chemical potential is greater by ψF, where F is the Faraday. $\mu^0(TP)$ is then defined as the chemical potential when $\theta = 0{\cdot}5$ and the electrostatic potential is zero.

For a proton
$$\mu_{\text{H}} = \mu^0_{\text{H}}(TP)_{\text{fibre}} + RT \ln \frac{\theta_{\text{H}}}{1-\theta_{\text{H}}} + \psi F, \tag{2}$$

and for a chloride ion, for example,

$$\mu_{\text{Cl}} = \mu^0_{\text{Cl}}(TP)_{\text{fibre}} + RT \ln \frac{\theta_{\text{Cl}}}{1-\theta_{\text{Cl}}} - \psi F. \tag{3}$$

The corresponding equations for these ions in a solution in equilibrium with the fibre, and referred to the same zero, are

$$\mu_{\text{H}} = \mu^0_{\text{H}}(TP)_{\text{soln.}} + RT \ln f_{\text{H}} . \text{H}, \tag{4}$$

and
$$\mu_{\text{Cl}} = \mu^0_{\text{Cl}}(TP)_{\text{soln}} + RT \ln f_{\text{Cl}} . \text{Cl}. \tag{5}$$

$\mu^0_{\text{H}}(TP)_{\text{soln.}}$ and $\mu^0_{\text{Cl}}(TP)_{\text{soln.}}$ are the chemical potentials of the hydrogen and chloride ions respectively when their activities are unity, while concentrations, in gram ions per litre, are represented by atomic symbols, and activity coefficients by f_{H} and f_{Cl}.

If the abbreviation
$$\Delta\mu_i = \mu^0_i(TP)_{\text{fibre}} - \mu^0_i(TP)_{\text{soln.}} \tag{6}$$

is used for the change in standard chemical potential of an ion of type i during adsorption, one finds by equating the chemical potential of the proton in the two phases, that

$$RT \ln \frac{\theta_{\text{H}}}{1-\theta_{\text{H}}} = -\Delta\mu_{\text{H}} + RT \ln \text{H} f_{\text{H}} - \psi F, \tag{7}$$

while for the chloride ion,

$$RT \ln \frac{\theta_{\text{Cl}}}{1-\theta_{\text{Cl}}} = -\Delta\mu_{\text{Cl}} + RT \ln \text{Cl} f_{\text{Cl}} + \psi F. \tag{8}$$

In treating $\Delta\mu_i$ as a constant with a value specific for each ion, we shall be assuming that the properties of a site such as its effective volume are approximately constant, and that variations in the term $\{(PV)_{\text{fibre}} - (PV)_{\text{soln.}}\}$ can be neglected.

ψ can be eliminated by adding equations (7) and (8), to give for the titration equation of hydrochloric acid,

$$RT \ln \frac{\theta_{\text{H}}\theta_{\text{Cl}}}{(1-\theta_{\text{H}})(1-\theta_{\text{Cl}})} = -(\Delta\mu_{\text{H}} + \Delta\mu_{\text{Cl}}) + RT \ln \text{H} f_{\text{H}}\, \text{Cl} f_{\text{Cl}}. \tag{9}$$

The unique importance of the hydrogen ion in practice is therefore not apparent in the titration equation, which is symmetrical with respect to anion and cation.

If there are the same number of acidic and basic sites,

$$\theta_{\text{H}} = \theta_{\text{Cl}}, \tag{10}$$

and for titration with pure acid there is the further equality,

$$\text{H} = \text{Cl}$$

leading to

$$\log \frac{\theta_{\text{H}}}{1-\theta_{\text{H}}} = -\text{pH} - \frac{\log_e}{2RT}\{\Delta\mu_{\text{H}} + \Delta\mu_{\text{Cl}}\}. \tag{11}$$

The base of the logarithms has been changed to allow the solution concentration to be expressed in terms of pH, and since the treatment already involves several approximations, we have set the activity coefficients equal to one.

Equation (11) is reasonably well satisfied by published data for the titration of wool keratin with hydrochloric acid. The experimental results of Steinhardt & Harris (1940) are compared with this theoretical relation in figure 1. In the calculation of θ_{H}, a maximum capacity of $8{\cdot}2 \times 10^{-4}$ g.mol. of hydrochloric acid per gram of wool has been assumed.

Titrations at constant ionic strength

When mixtures of ions of the same sign are present in solution, the fibre sites are shared according to the respective intrinsic affinities ($\Delta\mu_i$) and concentrations of the different ions. The effective fibre 'concentration' of any ion is no longer $\theta_i/(1-\theta_i)$ but is $\theta_i/(1-\Sigma\theta_i)$ (Fowler & Guggenheim 1939), where the summation includes all ions of the same sign as i.

If an alkali salt of the acid is added to the solution to enable titration to be carried out at constant ionic strength, equations analogous to (7) hold for both protons and alkali metal ions, e.g.

$$RT \ln \frac{\theta_{\text{H}}}{1-\theta_{\text{H}}-\theta_{\text{Na}}} = RT \ln \text{H} - \Delta\mu_{\text{H}} - \psi F, \tag{12}$$

and

$$RT \ln \frac{\theta_{\text{Na}}}{1-\theta_{\text{H}}-\theta_{\text{Na}}} = RT \ln \text{Na} - \Delta\mu_{\text{Na}} - \psi F \tag{13}$$

(again omitting activity coefficients, and using sodium ions as an example).

By subtraction,
$$\ln\frac{\theta_H}{\theta_{Na}} = \ln\frac{H}{Na} - \frac{1}{RT}\{\Delta\mu_H - \Delta\mu_{Na}\} \quad (14)$$

$(\Delta\mu_H - \Delta\mu_{Na})$ is strongly negative, and we can therefore neglect θ_{Na} in comparison with θ_H except when Na is very much greater than H.

Making use of this simplification, we find from equation (9) when sodium chloride is used to maintain a constant ionic strength, that

$$\log\frac{\theta_H}{1-\theta_H} = -\tfrac{1}{2}pH - \frac{\log e}{2RT}\{\Delta\mu_H + \Delta\mu_{Cl}\} + \tfrac{1}{2}\log Cl, \quad (15)$$

where Cl is now constant.

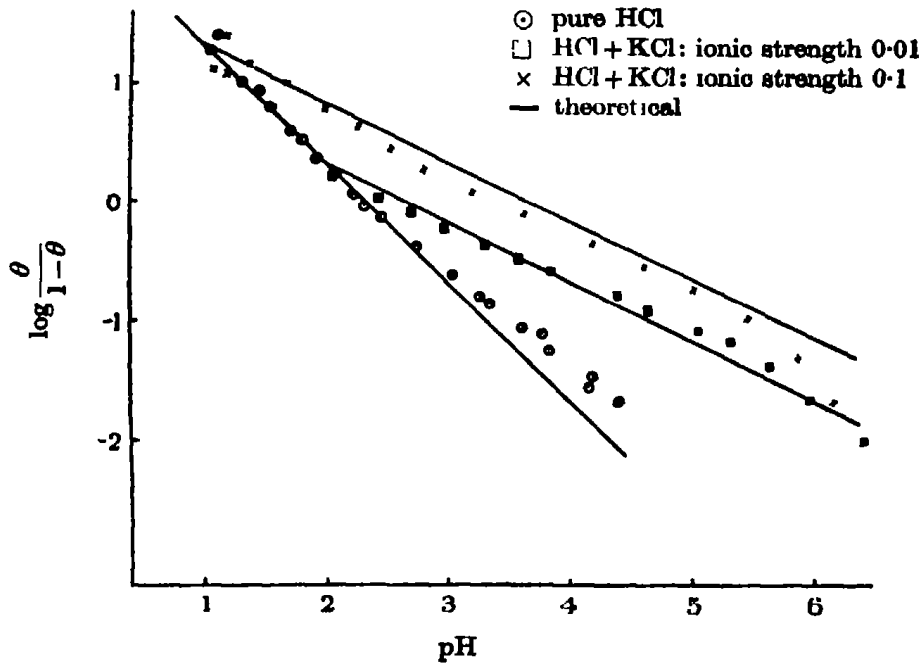


FIGURE 1. Titration of wool with HCl at 0° C by Steinhardt, Fugitt & Harris (1943).

Thus the greater the ionic strength, i.e. the greater Cl, the less the hydrogen-ion concentration needed for a given degree of fibre saturation. For instance, if the fibre is to be half-saturated

$$(pH)_{\theta=\frac{1}{2}} = \log Cl - \frac{\log e}{RT}(\Delta\mu_H + \Delta\mu_{Cl}). \quad (16)$$

Titrations carried out at constant ionic strength by Steinhardt (1942) and by Steinhardt & Harris (1940) confirm equation (15) (see figure 1), and their results for the dependence of the pH of half neutralization on ionic strength agree with equation (16) (figure 2).

The influence of the intrinsic affinity of the anion of an acid on the position of the titration curve with respect to the pH axis, is shown clearly by equation (11). For example, if the chloride ion is replaced by an ion of greater affinity (less positive $\Delta\mu_i$) the whole curve is moved to regions of higher pH.

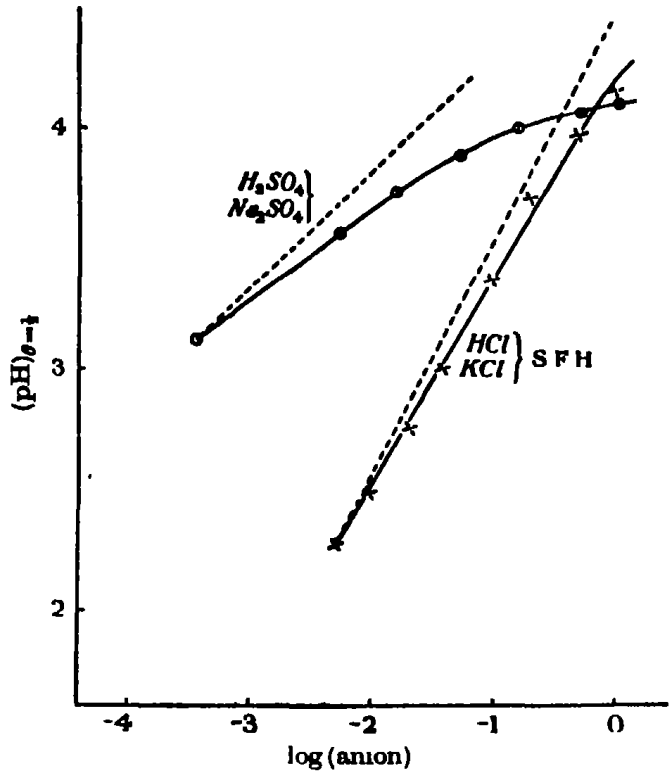


FIGURE 2. Titration of wool: pH for half-maximal combination.
----- theoretical slope.

RELATION TO OTHER THEORIES OF TITRATION

Very few quantitative theories for the titration of fibrous proteins have been published. The suggestion that the equations of Donnan (1911) developed for ionic equilibria across membranes should be applicable was made in 1925 by Speakman (1925), and two years later, Elöd & Silva (1927) interpreted their experiments on the swelling of wool in acids with the aid of these equations. They were influenced too much, however, by the analogous treatment of soluble proteins by Loeb (1922) in assuming that the water imbibed by proteins provided an ordinary aqueous environment for adsorbed ions. The present treatment differs in that it allows for interactions other than with the space charge or with water, and in taking account of the limited number of sites available for anions. The more general form of the equations of Donnan (Donnan & Guggenheim 1932) could, of course, be applied directly, if suitable activity coefficients were calculated for the fibre phase with the help of a model such as that in use here.

Recently, Steinhardt (1941, 1942) and Steinhardt & Harris (1940) have advanced very different ideas. It will be of help in presenting them if the physical meaning of the equations developed above is summarized.

It can be seen that in spite of the fibre potential being the only link between the adsorbed cations and anions, it does not appear in the final form of the titration equations. Even so, it provides the simplest key to the titration process. We may first consider titration with pure acid, as in the initial paragraphs of this paper. A very small quantity of acid added to an isoelectric fibre ensures that the fibre has a sufficient potential to adsorb anions even if they have almost no intrinsic affinity. The exact magnitude of this potential depends on the electrostatic energy needed to concentrate as many anions in the fibre as there are protons attached to carboxyl groups. This energy is decreased if one increases the concentration of anions in the external solution, or chooses an anion with more intrinsic affinity, but is increased if the desired concentration of anions in the fibre is increased. In the presence of pure acid, any lowering of pH is accompanied by an increased adsorption of acid, but no increase of fibre potential because the higher concentration of anions in the fibre is exactly counterbalanced by the higher anion concentration in the solution. The titration is therefore carried out with the fibre at a constant potential with respect to the solution, determined by the intrinsic affinity of the anion employed and the proton. If the anion contains groups with a high affinity for proteins, the fibre potential may even be zero or negative in acid solution.

The extent of the cleavage between this theory, and the treatment put forward by Steinhardt & Harris (1940) is immediately to be seen from their assumption, which we quote, that 'No recognition need be taken of differences in hydrogen-ion concentration inside and outside the fibre'. This may mean, they say, 'either that these concentrations are identical, or that they are related to one another in a way which is practically independent of the presence of other ions'.

As the basis of their treatment they assign four dissociation constants to the following equilibria:

$$W\,\mathrm{H}\,A \rightleftharpoons WA + \mathrm{H}, \quad K_{\mathrm{H}'}, \tag{17}$$

$$WA \rightleftharpoons W + A, \quad K_{A'}, \tag{18}$$

$$W\,\mathrm{H} \rightleftharpoons W + \mathrm{H}, \quad K_{\mathrm{H}}, \tag{19}$$

$$W\,\mathrm{H}\,A \rightleftharpoons W\,\mathrm{H} + A, \quad K_{A}, \tag{20}$$

whence

$$\frac{K_{\mathrm{H}'}}{K_{\mathrm{H}}} = \frac{K_{A}}{K_{A'}}. \tag{21}$$

In these equations W represents a salt-link such as —COO'NH_3^+— and WA, $W\,\mathrm{H}$ and $W\,\mathrm{H}\,A$ are its possible states of combination with an anion A' or proton H^+.

Drawing an analogy from the first and second dissociation constants of a dibasic acid, they assume that the tendency of H to dissociate from $W\,\mathrm{H}\,A$ is much less than its tendency to dissociate from $W\,\mathrm{H}$, i.e.

$$K_{\mathrm{H}} \gg K_{\mathrm{H}'}. \tag{22}$$

Similarly $$K_{A'} \gg K_A. \tag{23}$$

$pK_{H'}$ is given the value 4·2, and in one of the latest papers (Steinhardt 1942) a value of 100 is adopted for the ratio $(K_H/K_{H'})^{\frac{1}{2}}$.

They define the fraction of the fibre combined with acid by

$$\frac{WHA+WH}{WHA+WH+WA+W} \quad \text{or} \quad \frac{WHA+WA}{WHA+WH+WA+W}, \tag{24}$$

according to whether $$WH > \quad \text{or} \quad < WA. \tag{25}$$

In the first case the rest of the anions required for electrical neutrality are found in an adsorption layer round the fibre, and in the second case, protons form an external layer.

By substituting from equations (17) to (20) into (24), the fraction of the fibre combined can be related to the pH and anion concentration of the solution. It is then found that a plot such as that in figure 1 leads to a titration curve with a slope of about -2 instead of -1 for pure acid, and a slope of -1 instead of $-\frac{1}{2}$ for titration in the presence of a constant concentration of anions. To correct for this, Steinhardt (1942) proposes the introduction of all activities and constants as square roots, and suggests that this is necessary because the state of ionization of the fibre as a whole has not been taken into account.

In deciding the physical meaning of this theory, it is useful to consider three particular cases, namely,

(1) Titration with an acid of which the anion has almost no intrinsic affinity for proteins.

(2) Titration with an acid having an anion of very high affinity.

(3) Titration with an acid in which the anion has the same intrinsic affinity as the proton.

In the first case WA is negligible, and as a consequence (since

$$WHA = H \times WA/K_{H'}),$$

WHA must also be negligible. This leaves WH responsible for the whole of the adsorption, and requires practically all the anions to be in an adsorption layer round the fibre according to the postulates of the theory. As was pointed out earlier on, this could only be possible energetically if the protein unit were of the size of a soluble protein molecule. The second case is analogous to the first, with the position of the anions and protons reversed. In the third case, all the acid, for practical purposes, is present as ion-pairs, WHA, by the operation of relations (22) and (23). These ion-pairs function as undissociated molecules of acid with their chemical potential given by a relation similar to that for ions, i.e.

$$\mu_{HA} = \mu^0_{HA}(TP)_{\text{fibre}} + RT \ln \frac{\theta_{HA}}{1-\theta_{HA}}. \tag{26}$$

For equilibrium with the solution

$$(\mu_{\mathrm{H}A})_{\text{fibre}} = (\mu_{\mathrm{H}} + \mu_A)_{\text{soln.}}, \tag{27}$$

and therefore

$$\ln \frac{\theta_{\mathrm{H}A}}{1-\theta_{\mathrm{H}A}} = \ln \mathrm{H} + \ln A + \frac{(\mu^0_A)_{\text{soln.}} + (\mu^0_{\mathrm{H}})_{\text{soln.}} - (\mu^0_{\mathrm{H}A})_{\text{fibre}}}{RT}, \tag{28}$$

or

$$\log \frac{\theta_{\mathrm{H}A}}{1-\theta_{\mathrm{H}A}} = -2\mathrm{pH} + \frac{(\mu^0_{\mathrm{H}} + \mu^0_A)_{\text{soln.}} - (\mu^0_{\mathrm{H}A})_{\text{fibre}}}{2{\cdot}30RT}. \tag{29}$$

Equation (29) is the analogue of equation (11), and shows that the slope of -2 for the plot of $\log\theta/(1-\theta)$ instead of -1 does not arise through neglecting the charge on the fibre (which is uncharged during titration with an acid having properties that include it in class three), but is due to the assumption that adsorbed anions and protons are associated in the fibre.

The energy of association responsible for the difference between $K_{\mathrm{H}'}$ and K_{H} (and between K_A and $K_{A'}$) is numerically equal to $RT\ln K_{\mathrm{H}}/K_{\mathrm{H}'}$, which amounts to some 5 or 6 k.cal. if their value for the ratio of the constants is taken. This value seems far too high, since it arises from the interaction of a hydrated anion and a proton combined with a carboxyl ion, and we believe that the experimental evidence indicates that it is actually too small to be unambiguously detected. A different kind of association within the fibre is possible, however, and is found when the anion of a weak acid is adsorbed. Superimposed on the normal adsorption of anions and protons, one then finds an amount of the undissociated acid adsorbed proportional to the concentration of the undissociated acid in the external solution (Steinhardt *et al.* 1943). The dissociation constant of the adsorbed acid will not, of course, be the same as for the acid in aqueous solution.

Titration with polybasic acids

The theory just described has not been extended by its authors to include polyvalent ions, but these fit fairly easily into the scheme now being proposed, and it will be shown in a later paper that the latter can give meaning to a number of long-standing experiments with dyes.

In general, if an ion of charge z is adsorbed, the electrostatic part of its chemical potential in the fibre phase is $z\epsilon\psi$. The configurational part is more difficult to enumerate, for it cannot be said offhand how many positive sites a single polyvalent ion can occupy, or to what extent that number depends upon the size and shape of the ion. Until contrary evidence is obtained, it will be assumed that each ion occupies only one positive site. Then for an ion of type i, and valency z, electroneutrality requires that

$$\theta_i = \frac{\theta_{\mathrm{H}}}{z}, \tag{30}$$

where θ_i is the fraction of positive sites occupied.

For the adsorption of these anions from their solution of concentration c_i we can therefore write (cf. equation (8))

$$RT\ln\frac{\theta_i}{1-\theta_i} = RT\ln\frac{\theta_H}{z-\theta_H} = RT\ln c_i - \Delta\mu_i + zF\psi, \tag{31}$$

from which ψ may be eliminated using equation (7) to give

$$\ln\left(\frac{\theta_H}{1-\theta_H}\right)\left(\frac{\theta_H}{z-\theta_H}\right)^{1/z} = \frac{z+1}{z}\ln H - \frac{1}{RT}\left\{\Delta\mu_H + \frac{\Delta\mu_i}{z}\right\} - \frac{1}{z}\ln z \tag{32}$$

for the titration curve of pure acid in which

$$H = zc_i. \tag{33}$$

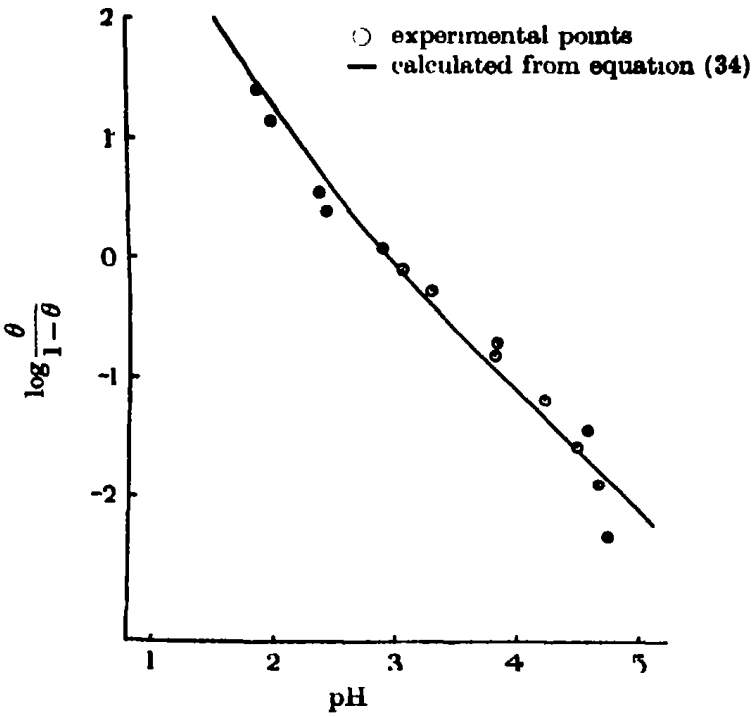


FIGURE 3. Titration of wool with H_2SO_4 by Speakman & Stott (1935).

This equation may be put into a form that resembles (11), thus

$$\log\frac{\theta_H}{1-\theta_H} = -\mathrm{pH} - \frac{z\log_e}{(z+1)RT}\left\{\Delta\mu_H + \frac{\Delta\mu_i}{z}\right\} + \frac{1}{z+1}\log\frac{1-\theta_H/z}{1-\theta_H}, \tag{34}$$

and differs from equation (11) essentially in the presence of a term which arises from the lower degree of saturation of the anion sites. Data obtained by Speakman & Stott (1935) for the titration of wool keratin with sulphuric acid have been compared in figure 3 with the curve resulting from equation (34) when z is put equal to 2.

The pH necessary for a given degree of saturation of the fibre increases as before if the anion concentration is increased, but according to a different law. Neglecting the adsorption of sodium ions, and letting c_i and H vary independently by adding the sodium salt of the acid, we find for the titration equation from (7) and (31) that

$$\ln\left(\frac{\theta_H}{1-\theta_H}\right)^z\left(\frac{\theta_H}{z-\theta_H}\right) = \ln c_i + z\ln H - \frac{(\Delta\mu_i + z\Delta\mu_H)}{RT}. \quad (35)$$

Thus if θ_H is kept constant,

$$(pH)_{\theta_H} = \frac{1}{z}\log c_i + \text{constant}. \quad (36)$$

An experiment was carried out in which samples of a specimen of wool containing $0{\cdot}41 \times 10^{-3}$ g.mol. of hydrogen ion per gram were washed with the minimum quantities of sodium sulphate solution of various strengths needed to ensure that each sample was in contact with a solution of known strength. The pH of each solution was then measured with a glass electrode. Negligible quantities of acid were removed during the washing and true equilibrium appeared to be attained almost immediately. From figure 2, in which the results are given, it can be seen that the slope of the plot of pH against log (SO_4) is initially 0·38 instead of the value 0·5 which one would predict from equation (36) At high concentrations of sulphate ion, the neglect of activity coefficients is serious, and sodium ions must begin to compete appreciably with protons for carboxyl sites.

We are indebted to Messrs I.C.I. (Dyestuffs) Ltd. for a grant to one of us (G. A. G.) which enabled this investigation to be carried out.

References

Adams, E. Q. 1916 *J. Amer. Chem Soc.* **38**, 1503–1510.
Abramson, H. A., Moyer, L. S. & Gorin, M H. 1942 *Electrophoresis of proteins.* N Y.: Reinhold.
Bjerrum, N. 1923 *Z. phys. Chem.* **104**, 147–173.
Cannan, R. K., Palmer, A. H. & Kibrick, C. A. 1941 *Ann. N.Y. Acad. Sci.* **41**, 243–266.
Cannan, R. K. 1942 *Chem. Rev.* **30**, 395–412.
Cannan, R. K., Palmer, A. H. & Kibrick, A. C. 1942 *J. Biol. Chem* **142**, 803–822.
Debye, P. & Huckel, E. 1923 *Phys. Z.* **24**, 185–206.
Donnan, F. G. 1911 *Z. Elektrochem.* **17**, 572.
Donnan, F. G. & Guggenheim, E. A. 1932 *Z. phys. Chem.* A, **162**, 346.
Elöd, E. & Silva, E. 1927 *Z. phys. Chem.* A, **137**, 142.
Fowler, R. H. & Guggenheim, E. A. 1939 *Statistical thermodynamics*, p. 426. Camb. Univ. Press.
Linderstrøm-Lang, K. 1924 *C.R. Lab. Carlsberg*, **15**, no. 7.
Loeb, J. 1922 *Proteins and the theory of colloidal behaviour.* N.Y.: McGraw Hill Book Co.
Speakman, J. B. 1925 *J. Soc. Dy. Col., Bradford*, **41**, 172.
Speakman, J. B. & Hirst, M. C. 1933 *Trans. Faraday Soc.* **29**, 148–172.
Speakman, J. B. & Stott, E. 1935 *Trans. Faraday Soc.* **31**, 1425.
Steinhardt, J. 1941 *Ann. N.Y. Acad. Sci.* **41**, 287–320.
Steinhardt, J. 1942 *Bur. Stand. J. Res., Wash.*, **28**, 191–199.
Steinhardt, J., Fugitt, C. H. & Harris, M. 1943 *Bur. Stand. J. Res., Wash.*, **30**, 123.
Steinhardt, J. & Harris, M. 1940 *Bur. Stand. J. Res., Wash.*, **24**, 335–367.

Efflux through a circular orifice

By Stefan G. Bauer

(*Communicated by Eric K. Rideal, F.R.S.—Received* 18 *January* 1943.—*Revised* 15 *August* 1943)

[Plate 11]

The purpose of this research was the establishment of the law of efflux from a vessel under pressure through a suddenly opened orifice of comparatively large dimensions

In order to simplify the approach to the problem, it was decided to investigate the efflux of an incompressible fluid, e.g. water The methods of approach were based on analogies with electric circuit theory, and accordingly an apparatus was built in which the equivalent of a c. experiments with resonant circuits could be performed. An orifice in the apparatus represented the inertia in the resonant system.

Acoustic theory gives the inertance of such an orifice for infinitesimal amplitudes In the experiments described, this inertance was measured at a wide range of finite amplitudes, by observations of phase shift between power input and displacement at the orifice By this means it was possible to determine the exact resonance frequencies of the hydraulic system in spite of its high damping. By this means, the effective mass or inertance of the fluid flowing through the orifice was deduced.

In the course of these experiments some interesting flow phenomena, to the author's knowledge not previously recorded, were observed and photographed There appear to be three distinct modes of flow, as compared with the two commonly observed in the steady state.

The contraction coefficient of the flow through the orifice was found to be unity for small amplitudes and to approach the steady state value for large amplitudes.

The differential equation of the efflux through a suddenly opened orifice is stated and its solution given. A numerical example is worked out with the constants derived from measurements and the results are shown by graphs They were checked by visual observation though a recording of the efflux experiment was not found possible It can be seen that a suction effect must follow a discharge. Further that this suction and the following oscillations are practically independent of the pressure in the vessel before the discharge, provided this pressure is above a certain value. These findings are of interest in connexion with the exhaust phenomena in internal combustion engines, in particular those working on the two-stroke cycle.

The problem of efflux of a fluid from a vessel through an orifice is usually treated by a step-by-step integration of a supposedly steady flow. When the rate of change of flow is small, this treatment provides a very good approximation. Recent developments, however, particularly in the field of two-stroke engines, have shown the inadequacy of this treatment of the efflux problem, which leaves out of account the inertia of the gas in the orifice. Hence a study of this aspect of the problem was undertaken. The problem is posed thus: What is the active inertia of a fluid flowing through a thin circular orifice in an infinite wall and what is the resistance opposing the flow? Further, does the latter differ from the resistance obtaining under conditions of steady flow?

In order to simplify the problem, it was decided to experiment with an incompressible fluid, i.e. water, and a suitable apparatus was accordingly built. Figure 1 shows diagrammatically the general arrangement. A pressure vessel of 12 in. diameter is divided horizontally into two halves by a diaphragm (D) to which different nozzles and orifices (O) can be secured. This diaphragm is made slightly

conical so as to bring the actual orifice into view through the 4 in. square plate-glass windows (W) let into the upper portion of the pressure vessel on two opposite sides. A 2 in. diameter window (L) is provided on the same level as, and half-way between, the two square ones for cross-illumination of the flow. The lower half of the vessel contains an inner cylinder (Q) 9 in. in diameter which makes up an annular space with the outside wall, closed on top and open at the bottom. This space can be filled with air through a valve (V) and its function is that of a spring in the oscillating system in which the mass is provided by the water in and around the orifice. Right

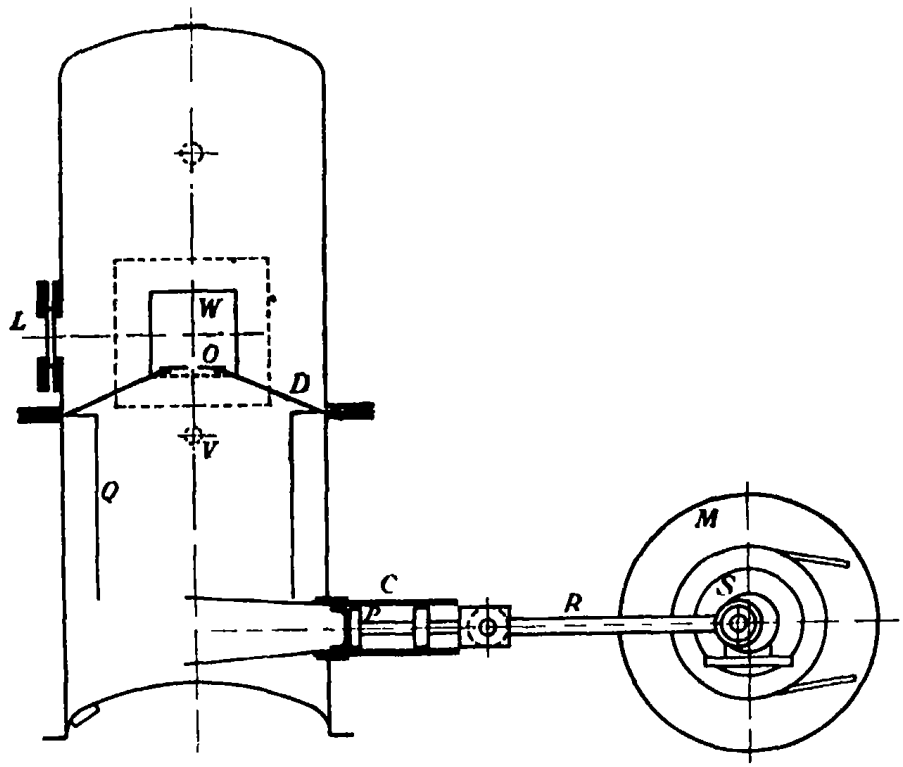


FIGURE 1

at the bottom of the apparatus is a horizontal cylinder (C) of 2 in. diameter in which a plunger (P) can be oscillated through a long connecting rod (R) by an adjustable throw crank (S) fixed to a heavy flywheel (M) and driven by a variable speed electric motor. Originally it had been intended to have air, compressed if necessary, above the water surface in the upper portion of the vessel only, to make forced oscillations of the water through the orifice by means of the plunger and to obtain a record of the periodic variations of pressure below the diaphragm. These records, it was thought, might subsequently be analysed into a Fourier series with a view to obtaining the values of inertia and resistance. These hopes were however disappointed in spite of a protracted struggle, because all manner of incidental pressure variations and oscillations completely covered up any inertia effects there may have been. The idea then presented itself of turning the whole thing into a free oscillating

system in which the mass represented by the inertia in the orifice could be accurately determined. The 'air spring' described above was added for this purpose. It will at once be apparent that this arrangement constitutes a simple Helmholtz resonator with water substituted for air in the orifice. Hence the first step was to find out to what extent and with what limitations the established theory applied to this particular case.

The resonance frequency of a Helmholtz resonator (Rayleigh 1878) is

$$f_r = \frac{1}{2\pi}\sqrt{\frac{2pr}{V\rho}} \tag{1}$$

for a circular orifice of negligible thickness and assuming isothermal compression and expansion of the enclosed air.

N sec.$^{-1}$	cycles per sec.	V cm.^{3}	volume of air spring
p kg. cm.$^{-2}$	pressure	ρ kg. sec^2 cm.$^{-4}$	density of fluid
r cm.	radius of orifice		

This formula is based on the assumption of uniform velocity over the whole area of the orifice and takes the inertia of the orifice to be that of a tube of the same diameter and the length (Rayleigh 1878)

$$l = \tfrac{1}{2}\pi r. \tag{2}$$

This is equivalent to twice the 'end-correction' for a pipe of length $l \to 0$ terminating in a flange.

The orifice finally adopted for the whole series of experiments was made of phosphor-bronze 0·3 mm. thick and soldered at the circumference to a brass ring which in turn was screwed to the diaphragm in the apparatus. The orifice was 2·231 cm. in diameter and had very carefully machined sharp edges. This is important because an orifice with rounded edges is equivalent to a thin and sharp-edged one of slightly larger diameter. For instance, an orifice plate, 3 mm thick, with edges rounded to 1·5 mm. radius and a diameter of 1·333 cm., was found to be equivalent to a thin and sharp-edged orifice of 1·485 cm. diameter.

With the orifice of 2·231 cm. diameter and an air-spring volume $V = 7700$ c.c., the resonance frequency for very small amplitudes was found experimentally as $f_r = 2{\cdot}7$ sec.$^{-1}$. Theoretically

$$f_r = \frac{1}{2\pi}\sqrt{\left(\frac{2\,.\,1}{7{\cdot}70\,.\,10^3}\,\frac{1{\cdot}27}{1{\cdot}02\,.\,10^{-6}}\right)} = 2{\cdot}71\,\text{sec.}^{-1}.$$

This is adequate proof, if proof were needed, of the validity of acoustical theory. The theory, however, considers small amplitudes only, and flow patterns assume quite a different character at larger amplitudes, as will be seen from the accompanying plates. The photographs were taken with cross-illumination of a plane through the axis of the orifice. They represent, as it were, a central cross-section

of the flow pattern. Figure 2 was taken at a small amplitude, about 0·5 mm., and exposed for half a dozen cycles or so. The slow random convection at the rate of 1 mm./sec. or thereabouts should be discounted. The amplitude is so small that the conditions approach those of infinitesimal amplitude postulated by acoustic theory. Accordingly it will be observed that the amplitude of oscillations increases from the centre of the orifice towards the edge without, of course, reaching the infinite velocity at the edge which should occur in the ideal fluid. The viscous boundary layer is responsible for that. It is quite possible that the inertance of the orifice is increased by the non-uniform velocity distribution but, owing to the high degree of damping, the resonance frequency cannot be determined to much less than 2 % error. The difference in theoretical frequency between uniform and outwardly increasing velocity distributions, as calculated by Rayleigh, is of the same order. Figure 3, which was taken under similar conditions as figure 2 but at an amplitude of about 1·5 mm., shows that this increase of velocity towards the edge disappears as soon as appreciable amplitudes are reached. Here the velocities are substantially uniform at any rate to very near the edge of the orifice, a fact which was confirmed by other observations and photographs not reproduced.

Another notable point is the movement of fluid from the edge towards the centre. This is due to the centrifugal force acting on those fluid particles which perform a semicircular movement from one face of the orifice to the other. This takes place however small the amplitude and its inevitable consequence is a slow movement axially away from both sides of the orifice. For very small amplitudes this movement consists of small forward and backward steps of each fluid element and friction dominates, so that each fluid particle stops and reverses when the displacement reverses. If the amplitude is increased beyond a definite critical limit, a sudden change takes place. The 'viscous structure' of the liquid suddenly breaks down and a slow, perfectly laminar jet develops along the axis of the orifice and away from it on both sides (see figure 4). These jets flow steadily at the rate of a few cm./sec. and the oscillations are merely superimposed on them in the neighbourhood of the orifice.

When the amplitude is increased beyond the point of the steady laminar regime, a different flow pattern makes its appearance. The instant the fluid begins to emerge through the orifice, a vortex ring starts from the orifice and is then followed by the jet proper. The jet is, as it were, capped by a vortex ring and looks somewhat like a mushroom rapidly growing out of the orifice. During the negative half cycle the orifice acts as an orthodox sink which cuts off the lower end of the outgoing jet. A static photograph gives only a very imperfect impression of this swiftly moving phenomenon (figure 5).

Whatever the amplitude and velocity the jet is always preceded by a vortex ring. Turbulence is most violent just behind the vortex ring and then settles down somewhat. The end of such a turbulent jet is shown in figure 6. The inward velocity near the orifice plane persists from the previous 'suction stroke'. In the orifice will be seen two specks in the process of reversing. No photograph of the

front end of the jet could be secured. Its appearance is very similar to the smoke cloud in front of a heavy gun a moment after it has been fired. The mechanism is presumably similar. It is this form of turbulent, intermittent jet which has frequently been observed in connexion with Helmholtz resonators.

For all the subsequent measurements and calculations, the customary acoustical units have been used (Olson 1940), and a short list of them together with their dimensions and mechanical and electrical equivalents is given below:

mechanical quantity	electrical quantity	acoustical quantity	unit	symbol	dimensions
force	e.m.f.	pressure	dynes/sq.cm.	P	$ML^{-1}T^{-2}$
displacement	charge or quantity	volume displacement	c.c.	X	L^3
velocity	current	volume current	c.c./sec.	$\dot{X}$	L^3T^{-1}
resistance	resistance	resistance	acoustical ohms	R	$ML^{-4}T^{-1}$
mass	inductance	inertance	g./cm.^{4}	M	ML^{-4}
compliance	capacitance	capacitance	c.c./(dynes/sq.cm.)	C	$M^{-1}L^4T^2$

The observation of resonance frequencies for different amplitudes was carried out in the following manner. A wooden float was placed on the upper free surface of the water and protected against the impinging jet from the orifice below. A glass tube fixed to the centre of the float was brought out clear of the apparatus. It carried at its topmost end a fine horizontal crosswire which could be viewed through a microscope containing an ocular scale. With the dimensions used the ratio of the displacements at the orifice and the float was 186·5 to 1, so quite a high order of magnification was needed to observe the smaller amplitudes. The curves in figure 7 show the actual volume displacements against frequency for different strokes of the driving plunger. It must be remembered that it is not the displacement but the volume current (velocity) which reaches a maximum at the point of resonance. The volume currents against frequency curves, calculated from the displacement curves, are shown in figure 8. The position of the current peaks thus found is not very accurate because a slight change of the slope of the displacement curve results in a considerable shift of the current peak. A different method of determining the resonance point independently was therefore resorted to.

The resonant frequency f_r of an acoustical system of one degree of freedom,

$$f_r = \frac{1}{2\pi\sqrt{(MC)}}, \tag{3}$$

and pressure and volume current are in phase. Hence pressure and displacement are 90° out of phase. In other words the frequency at which the displacement of the driving plunger and the displacement of the fluid are 90° out of phase, must be the resonant frequency. The arrangement sketched in figure 9 was thus used for determining the resonance frequencies for a series of different amplitudes.

A drum (D) with two slits (S) on opposite sides is fixed to the crankshaft (C) of the driving plunger (P) so that light from a strong electric bulb (B) passes through the slits at the inner and outer dead centre positions of the plunger. This light is collected by a lens (not shown) and directed by a mirror (M) on to the crosswire (I) and microscope (O). Extraneous light is kept out by screening.

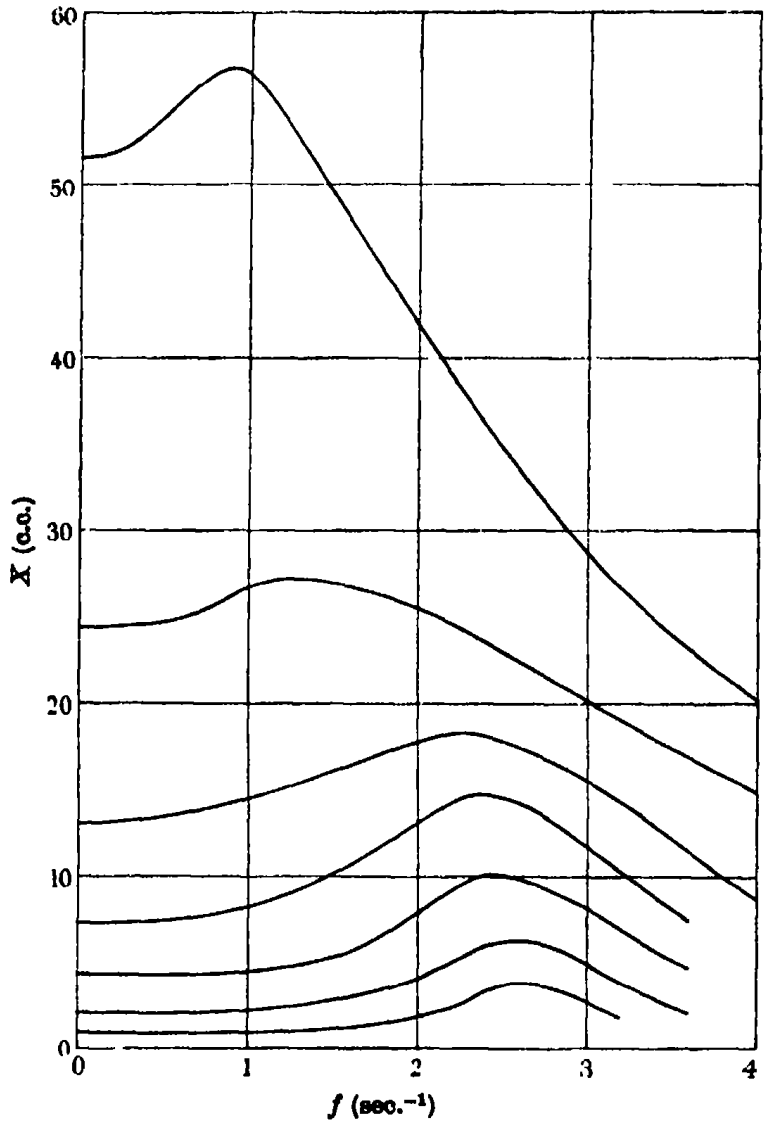


FIGURE 7

When the apparatus is set in motion the crosswire appears in the microscope twice every cycle as a dark line on the bright background. Though the time of the appearance is only very short—a fraction of a millisecond—the image is retained long enough in the observer's eye to form a fairly steady picture. Normally then, two distinct lines are observed, one produced on the up-stroke and the other on the down-stroke. As the phase difference between the driving plunger and the oscillating fluid



FIGURE 2



FIGURE 3



FIGURE 4



FIGURE 5



FIGURE 6

increases, the two lines in the field of the microscope approach one another and when the phase difference reaches 90°, i.e. the resonance point, the two lines coincide.

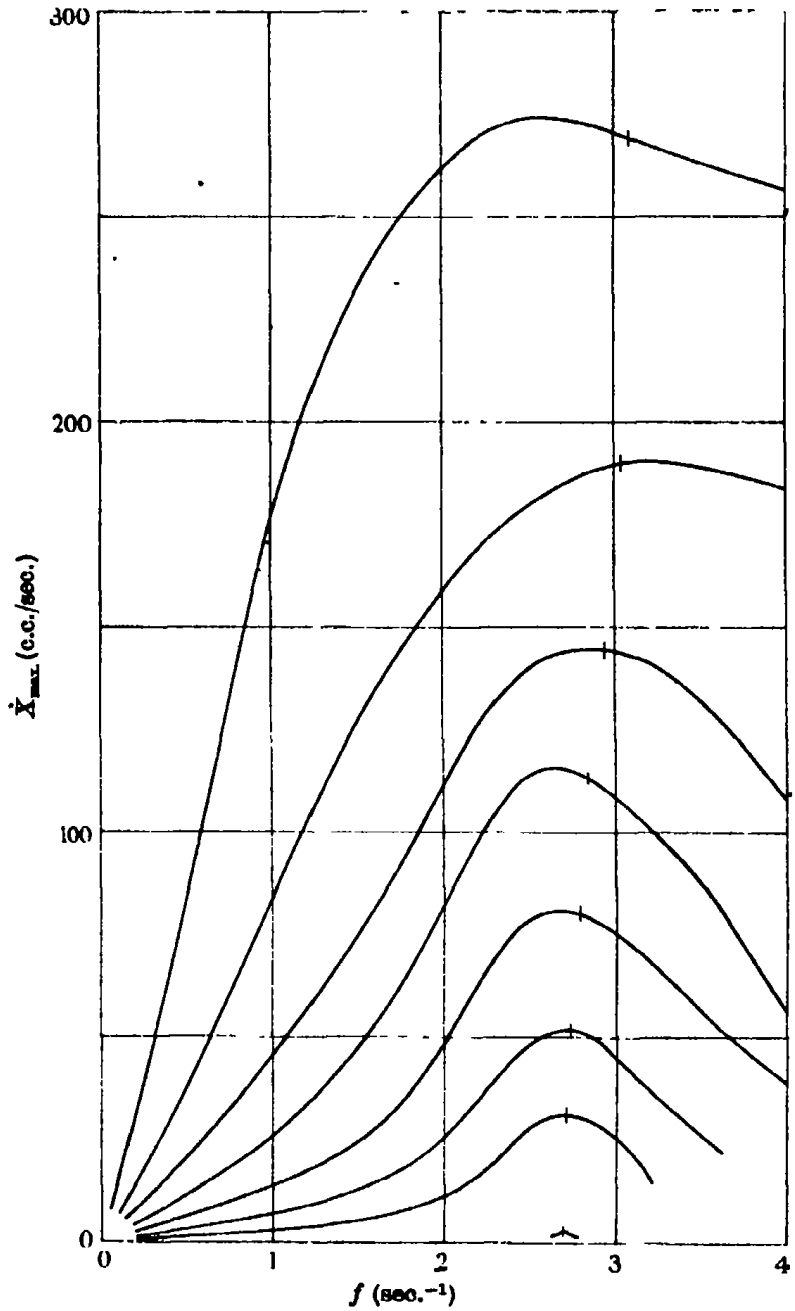


FIGURE 8

At the resonance frequency the fluid (and with it the crosswire) goes through its zero position at the moment when the driving piston is at its inner or outer dead centre.

Figure 10 shows the result of the experiment. The resonance frequency is shown plotted against the volume displacement of the driving piston. The increase of the

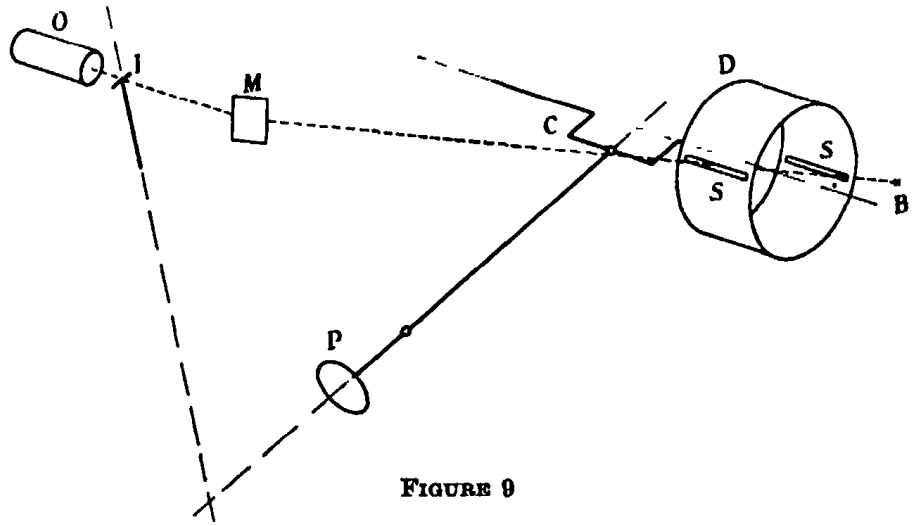


FIGURE 9

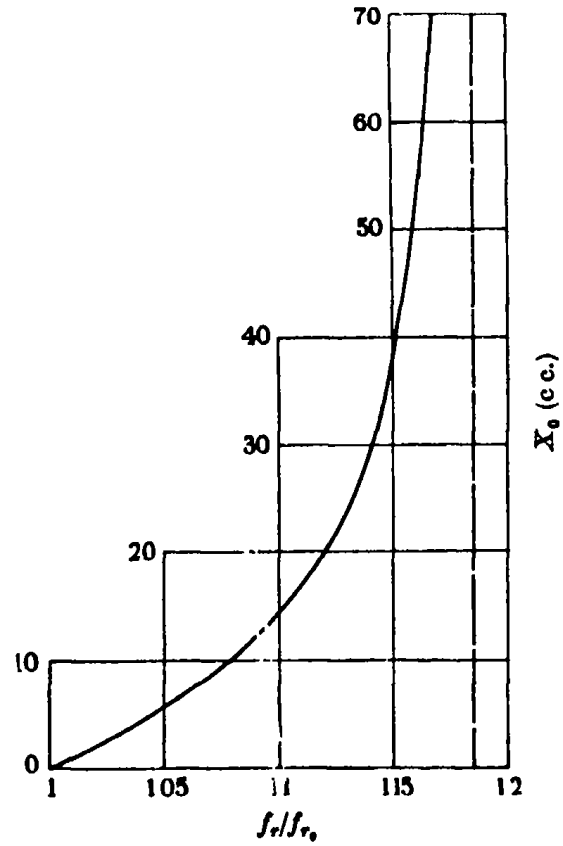


FIGURE 10

resonance frequency is explained by the fact that the pressure changes in the air spring (capacitance) are isothermal only for very small amplitudes. For adiabatic pressure changes, the frequency has to be multiplied by $\sqrt{\gamma} = 1{\cdot}185$ if γ for air is

taken as 1·4, and the curve is seen to approach this value asymptotically. Any small changes which may occur in the inertance of the orifice with increasing amplitudes are hidden by this property of the air spring.

From these results it seems fair to assume that theoretical formulae for the resonance frequency

$$f_r = \frac{1}{2\pi}\sqrt{\frac{2pr}{V\rho}} \tag{1}$$

under isothermal conditions and

$$f_r = \frac{1}{2\pi}\sqrt{\frac{2\gamma pr}{V\rho}} \tag{4}$$

under adiabatic conditions are substantially correct, irrespective of the amplitude of oscillation.

The inertance, capacitance and resistance of the 'resonator' may now be calculated. Let X be the volume displacement equal to displacement $x \times$ area of orifice A, then

$$X = Ax = 3{\cdot}91x.$$

The capacitance $$C = \frac{V}{\rho c^2} = 5{\cdot}910^{-3} = \frac{1}{169{\cdot}5},$$

where c is the velocity of sound.

According to (4) the adiabatic resonance frequency

$$f_r = 3{\cdot}19\,\text{sec.}^{-1},$$

whence with (3) we obtain the inertance

$$M = 0{\cdot}423,$$

and the equivalent oscillating mass m,

$$m = A^2M = 6{\cdot}45\,\text{g}.$$

At the resonance frequency, the fluid is at the zero displacement and maximum velocity when the driving piston is at its maximum displacement. Hence the pressure, responsible for the peak velocity, is that created by the full displacement of the driving plunger on the air spring. The displacement amplitude of the plunger is equal to the displacement amplitude in the orifice at zero frequency and is denoted by X_0 in the table below.

X_0 c.c.	f_r sec.$^{-1}$	p kg./m.^{2}	$v_{max.}$ cm./sec.
0·912	2·71	1·184	8·0
2·052	2·735	2·665	13·3
4·34	2·79	5·63	20·55
7·27	2·84	9·45	29·0
13·06	2·94	16·97	37·0
24·43	3·09	31·75	48·7
51·6	3·09	67·0	69·0

For small isothermal changes $$\frac{p}{p_0} = -\frac{X_0}{V}, \tag{5}$$

where p is the pressure due to displacement X_0, p_0 the static pressure, and V the volume of air.

In figure 11 are shown against $v^2_{max.}$ the pressure or resistance p also $v^2/2g$ and $v_1^2/2g$, where

$$v_1 = v/0{\cdot}62, \tag{6}$$

0·62 being the coefficient of contraction of the jet. v_1 is, in fact, the real velocity while v is the mean velocity over the orifice area. Considering the fact that all the kinetic energy of the oscillation is always lost with the jet, this result is to be expected. A contraction coefficient of 0·62 is in accordance with established experience for sharp edged orifices.

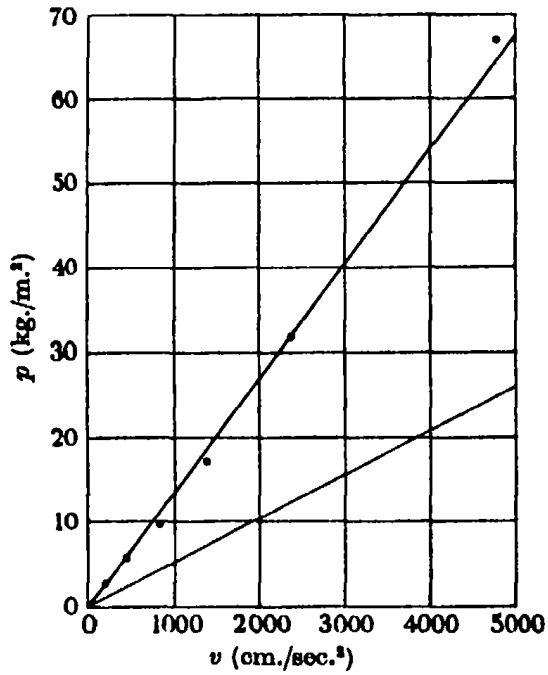


FIGURE 11

We can thus set

$$p = -\rho v^2/0{\cdot}62^2\,2g, \tag{7}$$

or in acoustical terms

$$p = 0{\cdot}085\dot{X}^2 \text{ in dynes/cm.}^2, \tag{8}$$

and

$$R = p/\dot{X}^2 = 0{\cdot}085. \tag{9}$$

The other constants of the system are

$$C = \frac{1}{169{\cdot}5}\ \frac{\text{cm.}^5}{\text{dynes}}, \qquad M = 0{\cdot}423\ \frac{\text{grams}}{\text{cm.}^4}.$$

We are now in a position to write down the differential equation of the efflux through the orifice, defined by the initial conditions:

$$t = 0, \quad X = X_0, \quad \dot{X} = 0,$$

$$M\ddot{X} + R\dot{X}^2 + \frac{X}{C} = 0. \tag{10}$$

The corresponding physical process is the sudden opening of the orifice with an initial displacement X_0 or an initial pressure on one side of the orifice. The solution of (10) may be arrived at by putting

$$\dot{X} = u, \tag{11}$$

whence

$$\ddot{X} = \frac{u\,du}{dX}, \tag{12}$$

and (10) becomes

$$Mu\frac{du}{dX} + Ru^2 + \frac{X}{C} = 0. \tag{13}$$

We now substitute

$$u^2 = 2q, \tag{14}$$

and get

$$M\frac{dq}{dX} + 2Rq + \frac{X}{C} = 0. \tag{15}$$

The solution of (15) has the form

$$q = e^{-(2R/M)X}\,y.$$

Differentiate this to

$$\frac{dq}{dX} = e^{-(2R/M)X}\left(\frac{dy}{dX} - \frac{2R}{M}y\right).$$

Hence

$$Me^{-(2R/M)X}\frac{dy}{dX} + \frac{X}{C} = 0,$$

and

$$y = -\frac{1}{MC}\int e^{(2R/M)X}\,X\,dX + y_0 = e^{(2R/M)X}\left(\frac{X}{2RC} - \frac{M}{4R^2C}\right) + y_0.$$

So the solution of (15) becomes

$$q = -\frac{X}{2RC} + \frac{M}{4R^2C} + y_0 e^{-(2R/M)X},$$

and the solution of (13)

$$u^2 = -\frac{X}{RC} + \frac{M}{2R^2C} + 2y_0 e^{-(2R/M)X}.$$

With y_0 determined from the initial conditions we arrive at the solution of (10) in terms of $\dot{X}$ and X,

$$\dot{X}^2 = \frac{1}{RC}\left[-X + \frac{M}{2R} - e^{-(2R/M)(X_0 - X)}\left(\frac{M}{2R} - X_0\right)\right] \tag{16}$$

wherein X has to change sign when $\dot{X}$ changes sign. Let for the sake of illustration $X_0 = -20$, then (16) reads with the above constants

$$\dot{X}^2 = 1995[-X + 2{\cdot}49 - 22{\cdot}49 \times e^{-0{\cdot}409(20-X)}].$$

The corresponding values of X and $\dot{X}$ are listed in the table below and shown in figure 12.

X	$\dot{X}$	X	$\dot{X}$	X	$\dot{X}$	X	$\dot{X}$
−20	0	−14	170·5	−4	113·6	2	31·2
−19	114·4	−12	165·0	−2	94·5	2·4	13·25
−18	145·3	− 8	155·5	0	70·5	2·48	4·0
−16	167·9	− 6	143·5	1	54·4	2·5	(−4·9)

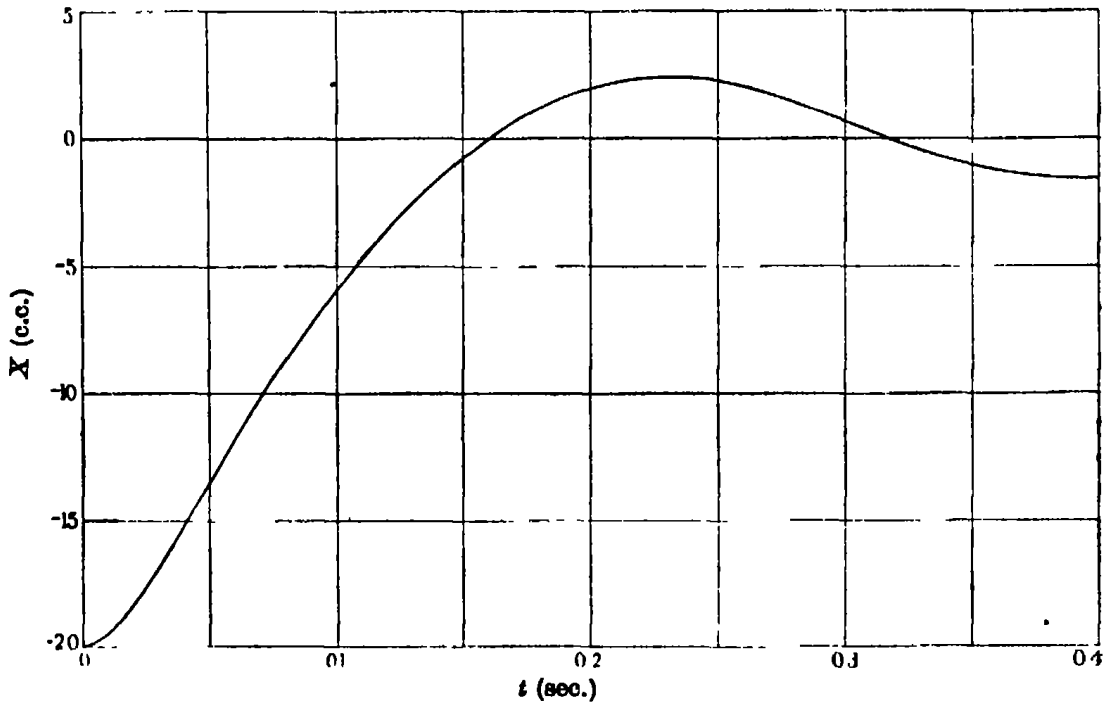


FIGURE 12

The volume current $\dot{X}$ returns to zero at $X = 2{\cdot}492$. This is the new X_0 for the next period of negative current where (16) becomes

$$\dot{X}^2 = 1995[X + 2{\cdot}49 - 4{\cdot}982 \times e^{-0{\cdot}409(2{\cdot}492-X)}],$$

and so forth. X and $\dot{X}$ for this period are listed below. They form the lower half of the spiral in figure 12.

X	$\dot{X}$	X	$\dot{X}$
2·492	0	0	37·20
2·4	13·55	−1	24·29
2	28·85	−1·5	5·74
1	39·60	−1·528	(−0·94)

The calculation could be carried on in the same manner but it is not really worth while because, for the following small amplitudes, viscosity would have to be taken into account.

The next and final step is a graphical integration. To obtain X in terms of t we integrate

$$\int \frac{1}{\dot{X}} dX = t + t_0. \tag{17}$$

This becomes difficult where $\dot{X}$ nears zero. At those points however the pressure and acceleration may be regarded as constant. This gives a parabolic displacement-time curve which is easily calculated. The approximation is better than the accuracy of a large-scale drawing. In figure 13 which shows the complete X-t curve, the indeterminate parts of the integration were filled in in this way. See also the table below.

The most noteworthy point about this efflux curve is the fact that the greatest secondary amplitude of the displacement in the opposite direction to the initial displacement is, but for a fraction of 1 %, independent of the initial displacement or pressure, provided, of course, that the latter is large enough. If we call this greatest secondary amplitude X_1 then we can write in good approximation

$$X_1 \simeq \frac{M}{2R},$$

provided we are dealing with a system where viscosity may be neglected. The other point is that, with the same limitation, this process of discharge always ends in a damped oscillation, unlike its electrical equivalent, i e. the discharge of a condenser through an impedance, which may become aperiodic. $M/2R$ in (18), it will be noted, is a function of the orifice radius r only. From (7) and (9) R is a function of ρ and r^{-4}. By definition

$$\text{inertance} = \frac{\text{mass}}{(\text{area})^2},$$

the mass m, $$m = \rho l \pi r^2,$$

where l is the length of the equivalent pipe. From (2) M is proportional to ρ and r^{-1}. Hence X_1 is proportional to r^3.

With a pipe in place of the orifice, the calculation would remain the same and the constants only would have to be altered. The inertance of a pipe of radius r and length l

$$M = \frac{\rho l}{\pi r^2}$$

against the circular orifice where $M = \frac{\rho}{2r}$.

Hence, with a pipe, the largest secondary amplitude would be proportional to length (including end corrections) and cross-sectional area.

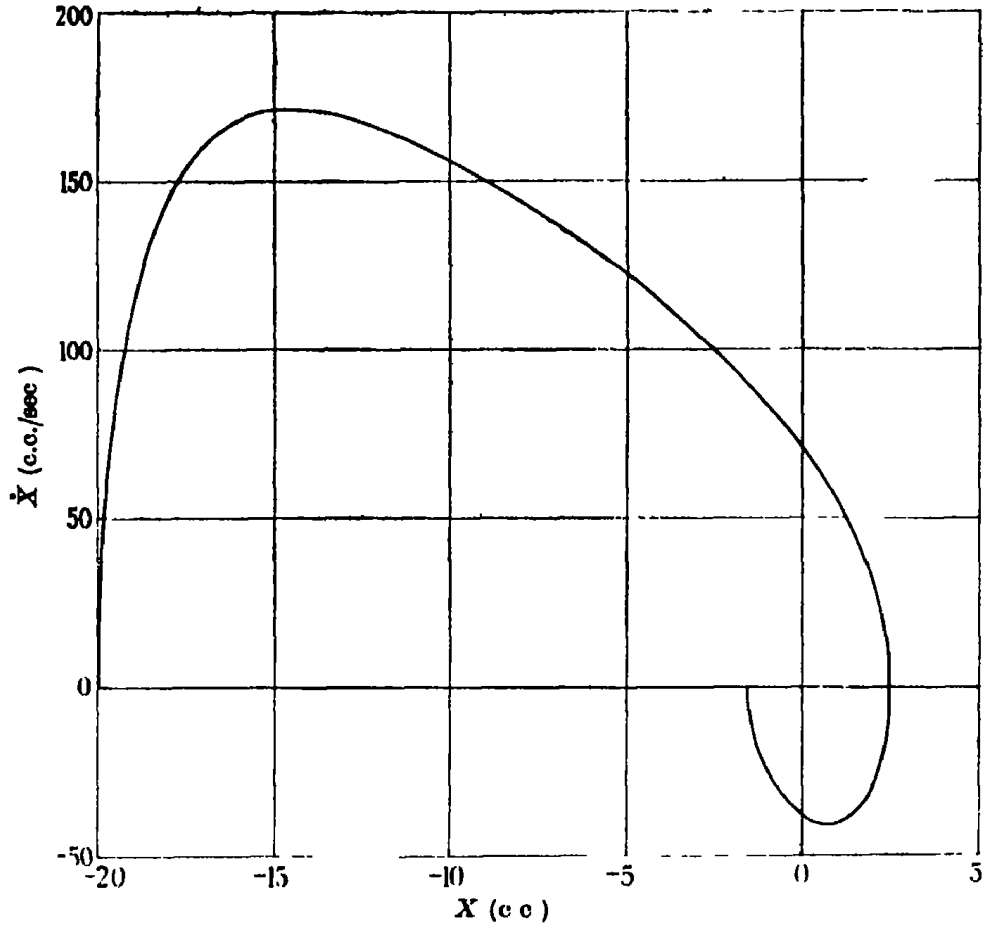


FIGURE 13

X (c.c.)	t (sec.)	X (c.c.)	t (sec.)	X (c.c.)	t (sec.)	X (c.c.)	t (sec.)
20	—	9	0 075850	− 2 492	—	0·2	0·07701
19·95	0	8	0·082670	− 2·4	0	0·4	0·08285
19·9	0·001515	7	0·089800	− 2·35	0·00327	0·6	0·08991
19·8	0 003640	6	0 097300	− 2·3	0·00594	0·8	0·09673
19·6	0·006675	5	0·105240	− 2·2	0 01050	0·9	0·10043
19·3	0·010095	4	0·113730	− 2 1	0·01448	1·0	0 10441
19·0	0·012945	3	0·122880	− 2·0	0·01808	1·1	0·10871
18·5	0·016995	2	0·132940	− 1·8	0·02448	1·2	0·11344
18·0	0·020580	1	0·144190	− 1·6	0·03036	1·25	0·11600
17·5	0·023905	0·5	0·150390	− 1·4	0·03582	1·30	0·11880
17	0·027100	0	0·157090	− 1 2	0·04104	1·35	0·12202
16	0·033190	− 0·5	0·164465	− 1·0	0·04612	1·40	0·12587
15	0·039090	− 1	0·172915	− 0·5	0·05862	1·45	0·13100
14	0·044950	− 1·5	0·183015	0	0·07149	—	—
13	0·050850	− 2	0·196515				
12	0·056840	− 2·2	0·203915				
11	0·062970	− 2·3	0·208565				
10	0·096290	− 2·4	0·214455				

Though the secondary volume displacement is in the first approximation a function of the orifice radius only, the efflux time is also a function of the fluid density, as can be inferred from equation (16). Thus the period, during which a depression exists in the experimental vessel after an efflux of water (see figure 13) would be more than a thousand times longer than the period of depression in a similar experiment carried out with air. Such comparisons, on a basis of density, are however only legitimate near zero amplitude where the compressibility of the gas can be neglected.

It was not found possible to obtain a complete experimental record of the discharge process corresponding to figure 13 but visual observation confirmed the general character of the curve. The time from the opening to the greatest negative amplitude was found to be about $\frac{1}{4}$ sec. The negative amplitude corresponded to the calculated figure as well as it could be observed in the very short time. Different initial displacement appeared to result in approximately the same peak negative displacements. The final oscillations died down more rapidly than figure 13 would seem to indicate. No doubt this is due to the viscosity of the water.

I wish to acknowledge my indebtedness to Professor C. E. Inglis, F.R S., for kind permission to carry out experiments at the Cambridge Engineering Laboratory and to Dr Derbyshire for his mathematical assistance

References

Olson 1940 *Elements of acoustical engineering.*
Rayleigh 1878 *Theory of sound*, **2**.

Asymptotic formulae relating to the physical theory of crystals

By Walter Ledermann

Department of Mathematics, The University, St Andrews

(*Communicated by M. Born, F.R.S.—Received 9 July* 1943)

In the physical theory of crystals great formal difficulties are encountered when the exact shape of the crystal is to be taken into account. Certain methods of approximation have therefore been developed and successfully used by several authors. However, as the validity of these methods was recently questioned by Sir C V. Raman, a more rigorous examination of the problem had to be undertaken.

It is found that the old procedure is fully justified provided the number of boundary particles is small compared with the total number of particles in the crystal. In particular, it is shown that lattice sums may in general be replaced by the corresponding infinite series, and that the distribution of frequencies follows with sufficient accuracy Born's law for cyclic crystals.

Upper bounds are obtained for the errors caused by these approximations.

Introduction

The problems treated in this paper were suggested by Professor M. Born. They are concerned with an examination of certain methods of approximation which have for some time been used in the physical theory of crystals and have generally been regarded as sound, though not rigorously proved.

However, Sir C. V. Raman and his collaborators (1941*a*, 1941*b*; cf. also Nagendra Nath 1943) have recently questioned the validity of the hitherto accepted theory, as developed by Born and others, and to end the controversy it seems necessary to subject the disputed points to a more careful mathematical scrutiny.

Lattice dynamics considers a crystal as a mechanical system of N mass points, which represent the nuclei of the atoms. The potential energy is obtained by averaging the total energy of the electronic motion for a given position of the nuclei. The equilibrium position of these is assumed to be a regular structure, a lattice, and most problems of lattice dynamics are concerned with small vibrations about this equilibrium. In tackling these problems the theory exploits to great advantage the regularity with which the equilibrium positions are arranged in space; indeed, its success would have been even more complete, were it not for the fact that this regularity breaks down on the boundary of the crystal and, in certain cases, also in the interior where a rather large number of particles fails to conform to the general structure. I shall therefore distinguish between *internal* and *boundary* points according as these points are surrounded by a complete set of regular neighbours or not (§§ 3 and 8). No detailed assumptions will be made with regard to the shape of the crystal or the position of boundary points, except that *their number shall be small compared with* N.

Part I deals with certain finite sums (lattice sums) extended over the N particles of a lattice. On account of the very great magnitude of N, these sums are very nearly equal to the corresponding infinite series (i.e. when $N \to \infty$), which are assumed to be absolutely convergent. This would almost seem to be a mere restatement of the definition of convergence, but some care is required to show (§ 4) that, despite the irregularities within the lattice, the finite and infinite series always have a sufficiently great number of *consecutive* terms in common so that, after a proper choice of origin, the remaining terms may be neglected.

In the discussion of the frequency distribution (part II) no such transition to infinite lattices is permissible, since the problem has a meaning only for a finite number of degrees of freedom ($3N$ in the case of N particles). Owing to the presence of boundary particles a rigorous treatment of the frequency problem appears to be impracticable. Born therefore modified the true boundary in accordance with his *cyclic condition* (§ 9) so as to preserve throughout the symmetry of the structure and of the corresponding equations. A complete solution can be obtained for this fictitious crystal, and it is reasonable to argue that it does not essentially differ from that for the actual crystal. Yet it is against this cyclic condition and its consequences that Raman's criticism was mainly directed.

With the help of a general theorem concerning the stability of latent roots of a Hermitian matrix (§ 11) I shall show that Born's statement with regard to the frequency distribution of a vibrating crystal is fully justified, and that his conclusions can be put on a rigorous basis

Part I. Lattice sums

1. Suppose that the lattice consists of N mass points which are specified by N sets of integers

$$l = (l_1, l_2, l_3).$$

The radius vector of the point (l_1, l_2, l_3) is $l_1\mathbf{a}_1 + l_2\mathbf{a}_2 + l_3\mathbf{a}_3$ where $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ define the fundamental cell. The functions which occur in lattice physics are of the form

$$F_N(Q) = \sum_l f_l e^{2\pi i(lQ)}, \tag{1}$$

or

$$G_N(Q) = \frac{1}{N}\sum_l \sum_{l'} g_{l-l'} e^{2\pi i[(l-l')Q]}, \tag{2}$$

where $Q = (Q_1, Q_2, Q_3)$ denotes a set of three independent variables and where

$$(Ql) = l_1 Q_1 + l_2 Q_2 + l_3 Q_3.$$

The functions of type (2) depend only on the difference of the cell indices and are therefore independent of the origin. Both functions are periodic with respect to Q_1, Q_2, Q_3, and in view of the large number of terms in the sums on the right

sides of (1) and (2) it seems natural to compare them with the infinite triple Fourier series

$$F(Q) = F_\infty(Q) = \sum_{\alpha=-\infty}^{\infty} f_\alpha e^{2\pi i(\alpha Q)} \tag{3}$$

and

$$G(Q) = G_\infty(Q) = \sum_{\alpha=-\infty}^{\infty} g_\alpha e^{2\pi i(\alpha Q)}, \tag{4}$$

respectively, whose absolute and uniform convergence will be assumed; as before* $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and $(\alpha Q) = \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3$.

A formal connexion between $F_N(Q)$ and $F(Q)$ can be established by means of the *shape function*

$$Z_N(Q) = \sum_l e^{2\pi i(lQ)}, \tag{5}$$

which is itself of type (1). Let

$$z_\alpha = 1, \text{ when } \alpha = l, \quad \text{and} \quad z_\alpha = 0, \text{ when } \alpha \neq l,$$

so that

$$Z_N(Q) = \sum_{\alpha=-\infty}^{\infty} z_\alpha e^{2\pi i(\alpha Q)}.$$

Now consider the integral (where $dP = dP_1 dP_2 dP_3$)

$$\int_0^1 F(P)\, Z_N(Q-P)\, dP = \sum_\alpha \sum_\beta f_\alpha z_\beta e^{2\pi i(\beta Q)} \int_0^1 e^{2\pi i[(\alpha-\beta)P]} dP$$

$$= \sum_\alpha \sum_\beta f_\alpha z_\beta \delta_{\alpha\beta} e^{2\pi i(\beta Q)} = \sum_\alpha f_\alpha z_\alpha e^{2\pi i(\alpha Q)} = \sum_l f_l e^{2\pi i(Ql)},$$

i.e.

$$F_N(Q) = \int_0^1 F(P)\, Z_N(Q-P)\, dP \tag{6}$$

Following Ewald (1940) I shall call this integral the *fold* of $F(Q)$ and $Z_N(Q)$ and denote it by

$$\widehat{FZ}_N(Q).$$

Thus (6) can be written in the form

$$F_N(Q) = \widehat{FZ}_N(Q). \tag{7}$$

An analogous relation exists between the functions $G_N(Q)$ and $G(Q)$; it is based on the *Laue factor*

$$L_N(Q) = \frac{1}{N} |\, Z_N(Q)\,|^2 = \frac{1}{N} \sum_l \sum_{l'} e^{2\pi i[(l-l')Q]}, \tag{8}$$

which is itself a function of type (2), and a similar calculation yields the result

$$G_N(Q) = \widehat{GL}_N(Q). \tag{9}$$

* Suffixes l, l', ... will be used to denote summation over the N particles of the lattice, whilst Greek letters α, α', β, ... refer to summation over geometrical lattice points whether they are occupied by material particles or not.

2. The formulae (7) and (9) hold for any value of N and for any arrangement of the N lattice points in space; but their practical usefulness is restricted by the fact that the actual evaluation of the sums (5) and (8) presents great difficulties except in certain cases where the lattice is of a particularly simple shape. For example, if the lattice is a parallelepiped whose edges are parallel to those of the fundamental cell and contain $2n_1+1, 2n_2+1, 2n_3+1$ particles respectively (so that $N = \prod_{r=1}^{3}(2n_r+1)$) and if the centre of the parallelepiped be taken as the origin, then

$$Z_N(Q) = \prod_{r=1}^{3} \frac{\sin \pi(2n_r+1)\,Q_r}{\sin \pi Q_r}. \tag{10}$$

Shape functions for other cases have been calculated by v. Laue (1936), Patterson (1939) and others.

In physical applications the number N is as a rule very large indeed and for many problems it is sufficient to know the *asymptotic* behaviour of the functions involved. More precisely, if N be treated as a parameter, it is the case $N \to \infty$ which is of greatest interest. Nevertheless, the formulae are valid for any finite value of N, and upper bounds will be obtained for the differences

$$|F_N(Q) - F(Q)| \quad \text{and} \quad |G_N(Q) - G(Q)|.$$

3 It will be found necessary to stipulate that the aggregate of N cells which constitute the physical lattice should satisfy the following topological condition: a lattice point (l_1, l_2, l_3) is said to be an internal point if all its $3^3 - 1 = 26$ neighbours

$$(l_1+\sigma_1,\ l_2+\sigma_2,\ l_3+\sigma_3)$$

also belong to the lattice, where $\sigma_1, \sigma_2, \sigma_3$ independently take the values $-1, 0, 1$ with the exception of the set $\sigma_1 = \sigma_2 = \sigma_3 = 0$. Any other lattice point is called a *boundary point*.* The number of boundary points in the given lattice is denoted by ν; it is then supposed that the ratio

$$\frac{\nu}{N} \tag{11}$$

is a small number, i.e. roughly speaking, the measure of the surface is assumed to be small compared with the measure of the volume. This condition excludes needle-shaped and laminal crystals. On the other hand we shall admit multiply-connected lattices so that some boundary points may be due to holes or irregularities within the lattice, provided the internal surface of such sponge-like crystals does not become so great that the number of points on it is comparable with the total number of lattice points. The ratio (11) plays an important part in estimating the error terms of the asymptotic formulae.

* It must, however, be realized that the present definition of internal and boundary points depends on the choice of the fundamental vectors, which, as is well known, are not in general uniquely determined.

4. One of the consequences of the condition introduced in the preceding section is that the dimensions of the *maximum parallelepiped* which can be inscribed in the crystal tend to infinity as $N \to \infty$. More precisely, let $\Pi_r(h)$ be the parallelepiped containing all points $t_1\mathbf{a}_1 + t_2\mathbf{a}_2 + t_3\mathbf{a}_3$, where

$$h_j - r \leqslant t_j \leqslant h_j + r \quad (j = 1, 2, 3), \tag{12}$$

r being a positive integer and $h = (h_1, h_2, h_3)$ being a set of three integers. Let n be the greatest integer such that at least one $\Pi_n(h)$, with suitable h, can be inscribed in the lattice. It will be shown that

$$n \geqslant \frac{1}{2}\left(\sqrt[3]{\frac{N}{\nu}} - 3\right). \tag{13}$$

For imagine the whole of space to be filled with parallelepipeds of type Π_{n+1} no two of which overlap,* and suppose that exactly q of them contain at least one point of the lattice. Since by (12) there are $(2n+3)^3$ lattice points in each Π_{n+1}, it follows that

$$N \leqslant q(2n+3)^3.$$

On the other hand, each of the q parallelepipeds Π_{n+1} contains at least one boundary point, because its suffix exceeds the maximum number n. Hence

$$\nu \geqslant q \tag{14}$$

and therefore

$$\frac{N}{\nu} \leqslant (2n+3)^3 \tag{15}$$

which is equivalent to (13). Thus if the number $\sqrt[3]{(N/\nu)}$ were to increase indefinitely, so would the maximum parallelepiped which can be inscribed in the crystal.

5. Now return to the special lattice sums $Z_N(Q)$ and $L_N(Q)$ defined in (5) and (8). It will be shown that both these functions would become periodic 'δ-functions' (in Dirac's sense), if the parameter N were allowed to tend to infinity.

First consider $Z_N(Q)$. If $\quad F(Q) = \sum_{\alpha=-\infty}^{\infty} f_\alpha e^{2\pi i(\alpha Q)}$

be any absolutely and uniformly convergent Fourier series whatsoever, we have by (1) and (6)

$$F_N(Q) \equiv \sum_l f_l e^{2\pi i(lQ)} = \int_0^1 F(P)\, Z_N(Q-P)\, dP.$$

As was seen on p. 365, the topological properties of the lattice imply that a parallelepiped $\Pi_n(h)$ can be inscribed in the crystal where h is a suitable lattice point. If this point is taken as origin, the lattice sum $F_N(Q)$ may be split up† as follows:

$$F_N(Q) = \sum_{l \leqslant \Pi_n} f_l e^{2\pi i(lQ)} + \sum_{l > \Pi_n} f_l e^{2\pi i(lQ)}. \tag{16}$$

* That is, every lattice point in space is accounted for once and only once.

† It is hoped that no misunderstanding will be caused by the abbreviations for the limits of summation: thus $l \leqslant \Pi_n$ means summation over all lattice points in or on Π_n, and $l > \Pi_n$ refers to all lattice points outside Π_n.

It is convenient to arrange the terms in the triple series

$$\sum_{\alpha=-\infty}^{\infty} f_\alpha e^{2\pi i(\alpha Q)}$$

in such a way that the summation is carried out according to concentric shells surrounding the origin. The rth shell will be denoted by π_r and consists of the $(2r+1)^3-(2r-1)^3$ lattice points which lie on the surface of the parallelepiped $\Pi_r(0)$. Since the series converges absolutely and uniformly, the terms may be rearranged thus*

$$F(Q)=\sum_{\alpha=-\infty}^{\infty} f_\alpha e^{2\pi i(\alpha Q)}=\sum_{r=0}^{\infty}\sum_{\rho=\pi_r} f_\rho e^{2\pi i(\rho Q)}.$$

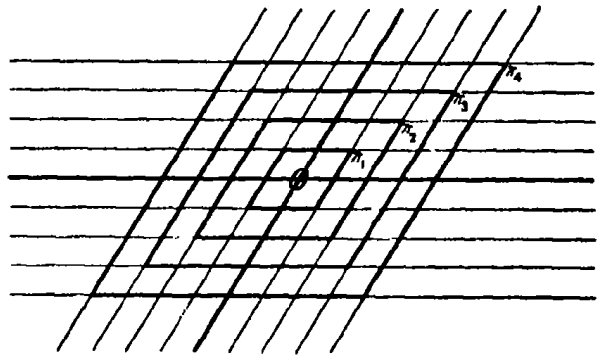


FIGURE 1. Illustrating the two-dimensional case.

In the first sum on the right-hand side of (16) the summation index l may be replaced by ρ, because the whole of Π_n belongs to the physical lattice. Hence

$$F_N(Q)=\sum_{r=0}^{n}\sum_{\rho=\pi_r} f_\rho e^{2\pi i(\rho Q)}+\sum_{l>\Pi_n} f_l e^{2\pi i(lQ)}. \tag{17}$$

On the other hand

$$F(Q)=\sum_{r=0}^{n}\sum_{\rho=\pi_r} f_\rho e^{2\pi i(\rho Q)}+\sum_{r=n+1}^{\infty}\sum_{\rho=\pi_r} f_\rho e^{2\pi i(\rho Q)}. \tag{18}$$

If

$$R_n=\sum_{r=n+1}^{\infty}\sum_{\rho=\pi_r}|f_\rho|,$$

it is evident that

$$\left|\sum_{l>\Pi_n} f_l e^{2\pi i(lQ)}\right|\leqslant R_n, \quad \left|\sum_{r=n+1}^{\infty}\sum_{\rho=\pi_r} f_\rho e^{2\pi i(\rho Q)}\right|\leqslant R_n,$$

and therefore on subtracting (18) from (17)

$$|F_N(Q)-F(Q)|\leqslant 2R_n. \tag{19}$$

The absolute convergence of the Fourier series $F(Q)$ implies that the remainder term R_n is small provided N/v and therefore n is large.

* The index ρ ranges over all lattice points on π_r.

Relation (19) can be written in the form

$$F_N(Q) = F(Q) + 2\theta(Q) R_n,$$

where

$$-1 \leqslant \theta(Q) \leqslant 1$$

for all values of Q. By expressing $F_N(Q)$ as an integral in accordance with (6) it is seen that

$$\int_0^1 F(P) Z_N(Q-P) dP = F(Q) + 2\theta(Q) R_n \tag{19'}$$

which may be compared with the property of the 'δ-function'

$$\int_0^1 F(P) \delta(Q-P) dP = F(Q).$$

The foregoing remarks are easily extended to the case of functions of two sets of variables Q and Q'. A finite sum of the form

$$F_N(Q, Q') = \sum_l \sum_{l'} f_{ll'} e^{2\pi i[(lQ)+(l'Q')]} = \sum_l \sum_{l'} u_{ll'} \tag{20}$$

is to be compared with the infinite series

$$F(Q, Q') = \sum_{\alpha=-\infty}^{\infty} \sum_{\alpha'=-\infty}^{\infty} u_{\alpha\alpha'}$$

which may be written*

$$F(Q, Q') = \sum_{\alpha\alpha' \leqslant \Pi_n \Pi_{n'}} u_{\alpha\alpha'} + \sum_{\alpha\alpha' > \Pi_n \Pi_{n'}} u_{\alpha\alpha'}, \tag{21}$$

where the modulus of the second double sum is not greater than

$$R_n = \sum_{\alpha\alpha' > \Pi_n \Pi_{n'}} |u_{\alpha\alpha'}| = \sum_{\alpha\alpha'} \sum_{\Pi_n \Pi_{n'}} |f_{\alpha\alpha'}|.$$

As before the remainder R_n tends to zero as n tends to infinity. Since all terms of the first double sum in (21) also occur in (20), it follows that

$$\left| \sum_l \sum_{l'} u_{ll'} - \sum_{\alpha=-\infty}^{\infty} \sum_{\alpha'=-\infty}^{\infty} u_{\alpha\alpha'} \right| \leqslant 2R_n,$$

i.e.

$$|F_N(Q, Q') - F(Q, Q')| \leqslant 2R_n, \tag{22}$$

or by analogy with (19')

$$\int_0^1 \int_0^1 F(P, P') Z_N(Q-P) Z_N(Q'-P') dP dP' = F(Q, Q') + 2\theta(Q, Q') R_n.$$

6. Consider now the second type of lattice sum, defined in (2), viz.

$$G_N(Q) = \frac{1}{N} \sum_l \sum_{l'} g_{l-l'} e^{2\pi i[(l-l')Q]} = \frac{1}{N} \sum_l \sum_{l'} v_{l-l'},$$

or

$$G_N(Q) = \frac{1}{N} \sum_{\alpha=-\infty}^{\infty} K(\alpha) v_\alpha,$$

* The six-dimensional parallelepiped $\Pi_n \Pi_n'$ now takes the place of Π_n.

where $K(\alpha)$ is the number of solutions of the equation

$$(l)-(l')=(\alpha). \tag{23}$$

When $(\alpha)=0$, this equation has evidently N solutions, i.e.

$$K(O)=N$$

and generally,
$$O\leqslant K(\alpha)\leqslant N. \tag{24}$$

It will be shown that for a certain range of values of α, $K(\alpha)$ is asymptotically equal to N.

Consider the case in which $(\alpha)=(\sigma^1)$ represents a fixed lattice point on the shell π_1. Equation (23) then has at least as many solutions as there are internal points in the physical lattice. For if (l') be any internal point, then by definition

$$(l')+(\sigma^1)=(l),$$

say, belongs to the lattice. As there are $N-\nu$ internal points, it follows that

$$K(\alpha)\geqslant N-\nu, \quad \text{if } (\alpha) \text{ lies on } \pi_1.$$

Next let $(\alpha)=(\sigma^2)$ correspond to a lattice point on π_2. Then

$$(\sigma^2)=(\phi')+(\psi'),$$

where (ϕ') and (ψ') represent certain points on π_1. There are at least $N-2\nu$ internal points (l') in the physical lattice which have the property that $(l')+(\phi')$ is also an internal lattice point, and consequently $(l')+(\phi')+(\psi')$ is a point of the lattice (internal or otherwise). Thus

$$K(\alpha)\geqslant N-2\nu, \quad \text{if } (\alpha) \text{ lies on } \pi_2.$$

Continuing in this way it is found that

$$K(\alpha)\geqslant N-r\nu, \quad \text{if } (\alpha) \text{ lies on } \pi_r \quad (r=1,2,\ldots)$$

and therefore
$$N\geqslant K(\alpha)\geqslant N-r\nu, \quad \text{if } (\alpha)\leqslant \Pi_r,$$

i.e.
$$1\geqslant \frac{K(\alpha)}{N}\geqslant 1-\frac{r\nu}{N}, \quad \text{if } (\alpha)\leqslant \Pi_r. \tag{25}$$

Let the integer n be defined by
$$n=\left[\sqrt{\frac{N}{\nu}}\right] \tag{26}$$

Then
$$1-\frac{n\nu}{N}\geqslant 1-\left(\sqrt{\frac{N}{\nu}}\right)\frac{\nu}{N}=1-\sqrt{\frac{\nu}{N}},$$

and by (25), when $r=n$,
$$1\geqslant \frac{K(\alpha)}{N}\geqslant 1-\sqrt{\frac{\nu}{N}},$$

i.e.
$$\frac{K(\alpha)}{N}=1-\theta(\alpha)\sqrt{\frac{\nu}{N}}, \quad \text{if } (\alpha)\leqslant \Pi_n, \tag{27}$$

where
$$O\leqslant \theta(\alpha)\leqslant 1.$$

Now split up the sum $G_N(Q)$ and write

$$G_N(Q) = \frac{1}{N}\sum_{\alpha=-\infty}^{\infty} K(\alpha)v_\alpha = \frac{1}{N}\sum_{\alpha\leqslant \Pi_n} K(\alpha)v_\alpha + \frac{1}{N}\sum_{\alpha>\Pi_n} K(\alpha)v_\alpha,$$

whence by (27)

$$G_N(Q) = \sum_{\alpha\leqslant \Pi_n} v_\alpha - \sqrt{\frac{\nu}{N}}\sum_{\alpha\leqslant \Pi_n}\theta(\alpha)v_\alpha + \frac{1}{N}\sum_{\alpha>\Pi_n} K(\alpha)v_\alpha.$$

On the other hand $$G(Q) = \sum_{\alpha=-\infty}^{\infty} v_\alpha = \sum_{\alpha\leqslant \Pi_n} v_\alpha + \sum_{\alpha>\Pi_n} v_\alpha,$$

and therefore

$$G_N(Q) - G(Q) = -\sum_{\alpha>\Pi_n} v_\alpha - \sqrt{\frac{\nu}{N}}\sum_{\alpha\leqslant \Pi_n}\theta(\alpha)v_\alpha + \frac{1}{N}\sum_{\alpha>\Pi_n} K(\alpha)v_\alpha. \tag{28}$$

Let $$R_h = \sum_{\alpha\geqslant \Pi_h} |v_\alpha| \quad (h = 0, 1, 2, \ldots).$$

The absolute convergence of the infinite series $G(Q)$ implies that R_h tends to zero as h tends to infinity.

Clearly,

$$\left|\sum_{\alpha>\Pi_n} v_\alpha\right| \leqslant R_{n+1},$$

$$\left|\sum_{\alpha\leqslant \Pi_n}\theta(\alpha)v_\alpha\right| \leqslant \sum_{\alpha\leqslant \Pi_n}|v_\alpha| \leqslant \sum_{\alpha=-\infty}^{\infty}|v_\alpha| = R_0,$$

$$\left|\frac{1}{N}\sum_{\alpha>\Pi_n} K(\alpha)v_\alpha\right| \leqslant \sum_{\alpha>\Pi_n}|v_\alpha| = R_{n+1},$$

since by (24)

$$0 \leqslant \frac{1}{N}K(\alpha) \leqslant 1.$$

Thus

$$|G_N(Q) - G(Q)| \leqslant 2R_{n+1} + \sqrt{\frac{\nu}{N}}R_0 = R'_n,$$

which is small, provided $\sqrt{(N/\nu)}$ and therefore n is large enough.

Expressing $G_N(Q)$ as an integral in accordance with (9), it is seen that

$$\int_0^1 G(P)L_N(Q-P)\,dP = G(Q) + \theta(Q)R'_n,$$

which may be compared with the symbolical equation

$$\int_0^1 G(P)\delta(Q-P)\,dP = G(Q).$$

7. As an example of the foregoing method consider the following lattice sum (Born 1942, p. 404):

$$D_N(Q, Q') = \sum_l \sum_{l'} e^{2\pi i(lQ)} D_{l-l'} e^{-2\pi i(l'Q')}, \tag{29}$$

where the D_α's are certain numerical constants. Let

$$F(Q) = \sum_\alpha f_\alpha e^{2\pi i(\alpha Q)} \tag{30}$$

be any Fourier series which converges absolutely and uniformly and let

$$D(Q) = \sum_\alpha D_\alpha e^{2\pi i(\alpha Q)}. \tag{31}$$

It is required to show that approximately

$$\int_0^1 D_N(Q, Q')\, F(Q')\, dQ' \sim F(Q)\, D(Q).$$

The integral is easily evaluated, and it is found that

$$\int_0^1 D_N(Q, Q')\, F(Q')\, dQ' = \sum_l \sum_{l'} f_{l'} D_{l-l'} e^{2\pi i(lQ)} \equiv \sum \sum_{l'} u_{ll'}$$

say. The corresponding infinite series is

$$\begin{aligned} \sum_\alpha \sum_{\alpha'} u_{\alpha\alpha'} &= \sum_\alpha \sum_{\alpha'} f_{\alpha'} D_{\alpha-\alpha'} e^{2\pi i(\alpha Q)} \\ &= \sum_\alpha \sum_{\alpha'} f_{\alpha'} D_{\alpha-\alpha'} e^{2\pi i(\alpha' Q)} e^{2\pi i[(\alpha-\alpha')Q]} \\ &= F(Q)\, D(Q), \end{aligned}$$

since $\alpha - \alpha'$ may be replaced by a single summation index which ranges from $-\infty$ to ∞. Thus on using the inequality (22), viz.

$$\left| \sum_l \sum_{l'} u_{ll'} - \sum_{\alpha=-\infty}^{\infty} \sum_{\alpha'=-\infty}^{\infty} u_{\alpha\alpha'} \right| \leqslant 2R_n,$$

it is found that $$\left| \int_0^1 D_N(Q, Q')\, F(Q')\, dQ' - F(Q)\, D(Q) \right| \leqslant 2R_n,$$

where $$R_n = \sum_{\alpha\alpha' \geqslant \Pi_n \Pi_{n'}} |\, u_{\alpha\alpha'}\,| = \sum_{\alpha\alpha' \geqslant \Pi_n \Pi_{n'}} |\, f_{\alpha'} D_{\alpha-\alpha'}\,|.$$

In the physical problem which gives rise to this lattice sum, the quantities D_β are negligible, unless the point (β) is fairly near the origin so that only comparatively few terms in

$$d = \sum_{\beta=-\infty}^{\infty} |\, D_\beta\,| \tag{32}$$

need be regarded as non-zero. Therefore, since

$$\sum_{\alpha\alpha' \geqslant \Pi_n \Pi_{n'}} |\, f_{\alpha'} D_{\alpha-\alpha'}\,| \leqslant \sum_{\alpha' \geqslant \Pi_n} |\, f_{\alpha'}\,| \sum_{\beta=-\infty}^{\infty} |\, D_\beta\,| = d \sum_{\alpha \geqslant \Pi_n} |\, f_\alpha\,|,$$

the result can be written in the form

$$\left| \int_0^1 D_N(Q, Q')\, F(Q')\, dQ' - F(Q)\, D(Q) \right| \leqslant 2d \sum_{\alpha \geqslant \Pi_n} |\, f_\alpha\,|. \tag{33}$$

Part II.* Density of frequencies

8. It was assumed in part I that the configuration of the N mass points was of a particularly simple nature; the structure was in fact what is known as a *simple lattice*. But the structure of a crystal is in general more complicated and must be regarded as the superposition of a certain number (s) of simple lattices. Thus in the usual notation (Born 1921) the radius vector of a typical point is written in the form

$$\mathbf{r}_k + l_1\mathbf{a}_1 + l_2\mathbf{a}_2 + l_3\mathbf{a}_3,$$

where $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k, \ldots, \mathbf{r}_s$ correspond to the particles of the base and where l_1, l_2, l_3 are integers specifying the cell. However, as far as this problem is concerned, no details of the geometrical configuration will be required, and I shall again denote the total number of particles by N.

On the other hand, it will be necessary to modify the definitions of internal and boundary points. The atomic interaction forces are such that each particle is perceptibly influenced only by its more immediate neighbours, which might be called *active neighbours*. Accordingly, a particle will be called an *internal* particle, if it possesses all its active neighbours (as required by the structure of the crystal). Should one or more active neighbours be missing, the particle will be termed a *boundary* particle. As before, the total number of boundary particles (including those on internal boundaries) will be denoted by ν, stress being laid on those cases in which the ratio ν/N is small.

9. When the method of small vibrations is applied to the N particles of the crystal, the equations of motion of the ith particle can be written in the form (Born 1921, p. 574)

$$m_i\ddot{x}_i = \sum_j (a_{ij}^{(x)}x_j + b_{ij}^{(x)}y_j + c_{ij}^{(x)}z_j) \quad (i, j = 1, 2, \ldots, N) \tag{34}$$

together with two similar equations for y_i and z_i, there being in all $3N$ equations of motion for the whole crystal The frequencies ω are obtained in the usual way by assuming that $\ddot{x} = -\omega^2 x$, etc , and then writing down the determinantal condition that the system of equations should have a non-trivial solution. A rigorous solution of the ensuing algebraical equation of degree $3N$ presents difficulties which in all but the simplest cases appear to be insurmountable On the other hand, what is required in the physical applications is not a knowledge of the individual frequencies, but only of their density on the real axis. Born (1921, p. 587) therefore devised the following approximate method: owing to the periodicity of the crystal the equations of motion have the same form for all internal particles, because each of these particles is surrounded by the same constellation of active neighbours. Only the motion of boundary particles is governed by different equations. If these irregularities are ignored, as if the crystal were *cyclically closed* and had no boundary particles at all, a perfectly regular system of equations would be obtained. The complete solution

* A brief summary of this work was published in a letter to *Nature* (1943).

for this modified system was obtained by Born, and it is natural to assume that these changes, which affect only a comparatively small number of equations, cannot materially influence the distribution of frequencies, which therefore will be nearly the same for the actual crystal.

It is the object of the following pages to confirm Born's statement and to obtain more precise conditions for its validity.

The argument is intimately related to a theorem of Cauchy (1829) on quadratic forms (see p. 377), but the method is independent of this theorem and could in fact easily be adapted to furnish a new proof of it.

10. I begin by proving the following (cf Pólya & Szegö 1925)

LEMMA: *Let $\mu_1 < \mu_2 < \ldots < \mu_m$ and η be real numbers and let $g_1 > 0, g_2 > 0, \ldots, g_m > 0$. Then the rational function*

$$\psi(x) \equiv \frac{f_1(x)}{\phi_1(x)} = (x-\eta) - \sum_{j=1}^{m} \frac{g_j}{x-\mu_j}, \tag{35}$$

where

$$\phi_1(x) = \prod_{j=1}^{m} (x-\mu_j) \tag{36}$$

has exactly $m+1$ real and distinct zeros $\rho_1 < \rho_2 < \ldots < \rho_{m+1}$ which are separated by the poles $\mu_1, \mu_2, \ldots, \mu_m$, i e.

$$\rho_1 < \mu_1 < \rho_2 < \mu_2 < \quad < \mu_m < \rho_{m+1}. \tag{37}$$

Proof It is readily seen that if ϵ is a sufficiently small real number

$$\psi(-\infty) < 0,\ \psi(\mu_1 - \epsilon) > 0,\ \psi(\mu_1 + \epsilon) < 0,\ \psi(\mu_2 - \epsilon) > 0,\ \psi(\mu_2 + \epsilon) < 0,\ \ldots,\ \psi(\infty) > 0,$$

which shows that $\psi(x)$ changes its sign at least $m+1$ times as x moves on the real axis from $-\infty$ to ∞. Hence one can locate at least $m+1$ roots which are separated by $\mu_1, \mu_2, \ldots, \mu_m$. On the other hand when $\psi(x)$ is written as the quotient of two prime polynomials $f_1(x)/\phi_1(x)$, the numerator is of degree $m+1$. Therefore $\psi(x)$ cannot have more than $m+1$ roots This proves the lemma

In what follows interest more particularly attaches to the distribution of roots on the x-axis Let

$$a \leqslant x \leqslant b$$

be any interval and let the number of roots of $f_1(x)$ and $\phi_1(x)$ be denoted by

$$P_1(a,b) \quad \text{and} \quad R_1(a,b)$$

respectively. Then it can easily be deduced from (37) that

$$|P_1(a,b) - R_1(a,b)| \leqslant 1, \tag{38}$$

no matter where the interval (a,b) is situated.

11. Consider the Hermitian matrix

$$A = \begin{bmatrix} c_{11} & \ldots & c_{1n} & h_1 \\ \ldots & \ldots & \ldots & \ldots \\ c_{n1} & \ldots & c_{nn} & h_n \\ \bar{h}_1 & \ldots & \bar{h}_n & \eta \end{bmatrix} = \begin{bmatrix} C & h \\ \bar{h}' & \eta \end{bmatrix}$$

of order $n+1$, where

$$C = \begin{bmatrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{n1} & \dots & c_{nn} \end{bmatrix}$$

is an Hermitian matrix of order n, and where

$$h = \{h_1, h_2, \dots, h_n\}$$

is a column vector and η is a real number. Let the characteristic polynomials of A and C be

$$f(x) = |xI - A| = \prod_{i=1}^{n+1} (x - \alpha_i)$$

and

$$\phi(x) = |xI - C| = \prod_{i=1}^{n} (x - \gamma_i)$$

respectively, where the α's and γ's are real numbers. There exists a unitary matrix U_0 of order n such that

$$\bar{U}_0' C U_0 = \begin{bmatrix} \gamma_1 & & & \\ & \gamma_2 & & \\ & & \ddots & \\ & & & \gamma_n \end{bmatrix} = \Gamma$$

say. Hence if the unitary matrix U of order $n+1$ is defined by

$$U = \begin{bmatrix} U_0 & 0 \\ 0 & 1 \end{bmatrix},$$

it readily follows that

$$\bar{U}' A U = \begin{bmatrix} \Gamma & e \\ \bar{e}' & \eta \end{bmatrix} = \begin{bmatrix} \gamma_1 & & & & e_1 \\ & \gamma_2 & & & e_2 \\ & & \ddots & & \vdots \\ & & & \gamma_n & e_n \\ \bar{e}_1 & \bar{e}_2 & \dots & \bar{e}_n & \eta \end{bmatrix}, \tag{39}$$

where

$$e = \bar{U}_0'.$$

Since the characteristic equation of $\bar{U}' A U$ is the same as that of A, $f(x)$ can be evaluated by expanding the characteristic determinant of (39) with respect to the last row and column (Aitken 1939), thus

$$f(x) = (x - \eta)\phi(x) - \sum_{i=1}^{n} (x - \gamma_1) \dots (x - \gamma_{i-1}) |e_i|^2 (x - \gamma_{i+1}) \dots (x - \gamma_n),$$

i.e.

$$\frac{f(x)}{\phi(x)} = (x - \eta) - \sum_{i=1}^{n} \frac{|e_i|^2}{x - \gamma_i}.$$

This formula holds for all cases including those in which some of the γ's are identical and some of the e's happen to be zero. On collecting terms which have the same denominator and omitting terms which are zero, it is seen that

$$\frac{f(x)}{\phi(x)} = (x - \eta) - \sum_{j=1}^{m} \frac{g_j}{x - \mu_j}, \tag{40}$$

where the μ's are distinct, each of them being equal to one of the γ's, and where the g's are positive. Hence all conditions of the lemma are fulfilled, and if the function (40) is reduced to its lowest denominator

$$\frac{f(x)}{\phi(x)} = \frac{f_1(x)}{\phi_1(x)} \tag{41}$$

say, it follows that, in the notation of (38), § 10,

$$|P_1(a,b) - R_1(a,b)| \leqslant 1. \tag{42}$$

Equation (41) implies the existence of a polynomial $h(x)$ such that

$$f(x) = h(x) f_1(x), \quad \phi(x) = h(x)\phi_1(x).$$

Therefore if the numbers of the roots of the functions

$$f(x), \quad \phi(x), \quad h(x)$$

in $a \leqslant x \leqslant b$ be denoted by $\quad P(a,b), \quad R(a,b), \quad S(a,b)$

respectively, then

$$P(a,b) = S(a,b) + P_1(a,b), \quad R(a,b) = S(a,b) + R_1(a,b),$$

whence by (42) $$|P(a,b) - R(a,b)| \leqslant 1. \tag{43}$$

It is on this relation between the roots of a Hermitian matrix A and one of its first minors C that the argument will be based Cauchy's theorem, which might easily have been deduced by the above method, goes a little further and states that the roots of A are separated by those of C.

Consider now a second Hermitian matrix of order $n+1$

$$B = \begin{bmatrix} c_{11} & \cdot & c_{1n} & k_1 \\ \cdots & \cdots & \cdot\cdot & \cdot \\ c_{n1} & \cdots & c_{nn} & k_n \\ \bar{k}_1 & \cdots & \bar{k}_n & \kappa \end{bmatrix} = \begin{bmatrix} C & k \\ \bar{k}' & \kappa \end{bmatrix},$$

which has the leading minor C in common with A (§ 11). If

$$Q(a,b)$$

denotes the number of latent roots of B in the interval, then by analogy with (43)

$$|Q(a,b) - R(a,b)| \leqslant 1,$$

and therefore $$|P(a,b) - Q(a,b)| \leqslant 2. \tag{44}$$

Since the latent roots of a matrix are not changed by a simultaneous permutation of its rows and columns, it is clear that the above result remains valid when the leading minor C is replaced by any other principal minor of order n.

If the Hermitian matrices are of order $n+r$ and have a principal minor of order n in common, but differ in the remaining r rows and columns, then by a repeated application of (44)

$$|P(a,b) - Q(a,b)| \leqslant 2r. \tag{45}$$

Thus the following theorem is proved:

If in a Hermitian matrix the elements of r rows and their corresponding columns are modified in any way whatever, provided only that the matrix remains Hermitian, then the number of latent roots which lie in any given interval cannot increase or decrease by more than 2r.

It should be noted that this property is independent of the order of the matrix.

12. I shall now resume the discussion of the frequencies of a vibrating crystal. By (34) the squares of the $3N$ frequencies are obtained by solving an algebraical equation of the form

$$|\omega^2 M - V| = 0, \tag{46}$$

where M is a diagonal matrix whose elements are $m_1, m_2, \ldots, m_N$ each occurring three times. The elements of $2V$ are the coefficients of the potential energy of the system, which is a positive-definite quadratic form in $3N$ variables, viz. the $3N$ components of displacement of the N particles. Contrary to the usual practice of specifying the particles by means of base and cell indices, I shall simply assume that the $3N$ components of displacement are numbered consecutively from 1 to $3N$, in such a way that the first $3(N-\nu)$ variables correspond to the internal particles and the last 3ν variables correspond to the ν boundary particles. Hence the last 3ν rows and columns of V represent those terms in the potential energy which are due to the interaction of boundary particles, either with themselves or with internal particles. I shall accordingly partition the matrix V thus

$$V = \begin{bmatrix} V_{ii} & V_{ib} \\ V_{bi} & V_{bb} \end{bmatrix} \begin{matrix} 3(N-\nu) \text{ rows} \\ 3\nu \text{ rows} \end{matrix},$$

where the suffixes i and b refer to internal and boundary particles respectively.

Equation (46) is equivalent to

$$|\omega^2 I - A| = 0, \tag{47}$$

where

$$A = M^{-\frac{1}{2}} V M^{-\frac{1}{2}} = \begin{bmatrix} A_{ii} & A_{ib} \\ A_{bi} & A_{bb} \end{bmatrix}$$

is a real symmetric matrix of order $3N$. The problem is therefore reduced to finding the latent roots of A. As has been pointed out, the rigorous solution of (47) is impracticable, but by modifying some of the terms in accordance with the cyclic condition one can obtain a matrix

$$B = \begin{bmatrix} A_{ii} & B_{ib} \\ B_{bi} & B_{bb} \end{bmatrix},$$

whose latent roots can be calculated. Since A and B differ only in their last 3ν rows and columns, the theorem of § 10 can be applied with the inference that

$$|P(a,b) - Q(a,b)| \leqslant 6\nu,$$

where the numbers of latent roots of A and B which lie in $a \leqslant x \leqslant b$ are denoted by $P(a,b)$ and $Q(a,b)$ respectively.

It is, however, more convenient to express the result in terms of the *density* of frequencies. Thus if

$$p(a,b) = P(a,b)/3N, \quad q(a,b) = Q(a,b)/3N$$

the above inequality takes the form

$$| p(a,b) - q(a,b) | \leqslant \frac{2\nu}{N}. \tag{48}$$

Hence, just as in part I, the conclusion is reached that the boundary particles have only a negligible influence on the density of frequencies provided the critical ratio ν/N is sufficiently small.

I should like to take this opportunity of expressing my sincere thanks to Professor Born for acquainting me with these problems and for the kind interest he has shown in my work.

References

Aitken, A. C. 1939 *Determinants and matrices*, p. 74. Edinburgh.
Born, M. 1921 Atomtheorie des festen Zustandes, *Encykl. Math. Wiss.* (5), **25**, 532.
Born, M. 1942 *Proc. Roy. Soc.* A, **180**, 397.
Born, M. & Ledermann, W. 1943 *Nature, Lond.*, **151**, 197.
Cauchy, A. L. 1829 *Exercises Math.* **4**, 153.
Ewald, P. P. 1940 *Proc. Phys. Soc.* **52**, 165.
v. Laue, M. 1936 *Ann. Phys., Lpz.*, (5), **26**, 55.
Nagendra Nath, N. S. 1943 *Nature, Lond.*, **151**, 196.
Patterson, A. L. 1939 *Phys. Rev.* (2), **56**, 972.
Pólya, G. & Szegö, G. 1925 *Aufg. Lehrs. Analysis* (Berlin), **2**, no. 26, 41.
Raman, C. V. & collaborators 1941*a* *Proc. Indian Acad. Sci.* A, **14**, 317.
Raman, C. V. & collaborators 1941*b* *Proc. Indian Acad. Sci.* A, **14**, 459.

The dissociation of an alloy of copper, iron and nickel
Further X-ray work

By Vera Daniel, Ph.D. and H. Lipson, D.Sc, *Cavendish Laboratory, Cambridge*

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 13 *September* 1943)

[Plate 12]

In a previous paper it was shown that X-ray evidence led to the conclusion that a series of intermediate structures is formed during the dissociation of the alloy Cu_4FeNi_3. Attempts to determine these structures failed because they led to X-ray intensities that did not agree with those observed, although their positions were adequately accounted for. In the present paper it is shown that the intensities are modified in a systematic way by extinction, and that, after introducing a correction for extinction, the theory advanced can account for all the data.

In the type of structure postulated there is a periodic variation of the lattice parameter. Concentration differences are set up in the alloy at regular intervals, while the coherence of the single-phase lattice remains. The wave-length of the periodicity is of the order of 10^{-6} cm.

On prolonged annealing at constant temperature the diffraction pattern changes gradually It is possible to follow the wave-length and the amplitude of the periodic variation as a function of time and temperature over a considerable interval. It has been found that the amplitude is independent of time at a given temperature and that the wave-length increases linearly with the logarithm of time The values of wave-length and amplitude as a function of time and temperature are reproducible within the limit of error, so that the measurements have a quantitative value. After longer times of annealing the diffraction pattern of the periodic state gradually changes into another pattern, which can be explained either by a periodic structure with a range of longer periods or by tetragonal lamellae

The bearing of the present observations on the kinetics of phase change is discussed.

Introduction

In a previous paper (Daniel & Lipson 1943) an account was given of an unusual diffraction effect shown by the alloy Cu_4FeNi_3 This alloy is single phase (face-centred cubic) at high temperatures and dissociates at lower temperatures into approximately equal amounts of two face-centred cubic phases, the dissociation is slow, and intermediate stages can be distinguished. The first stage of the dissociation gives a characteristic diffraction pattern in which, on a powder photograph, each line of the single-phase pattern is flanked by two side-bands. The position of these side-bands and also the evidence of oscillation photographs of coarse-grained specimens could be explained satisfactorily by the assumption of a periodic variation of the lattice parameter, the variation being in the direction of the cube edges. The intensities, however, did not fit in with the theory. This was thought to be so because the simple model chosen was too perfect. It now appears that the results can be explained even with this model, by the presence of a large amount of extinction.

It is not usual to associate large extinction effects with powder photographs, but it must be pointed out that in this case conditions are particularly favourable for its occurrence; in order to put the powder into a single-phase state it has to be so heat-treated that it always sinters and this means that the crystal grains must be larger than they usually are in powders.

Theory

The calculation of the structure factor for a one-dimensional grating with a periodic variation of the spacing shows that besides the orders of reflexion, h, for the undisturbed grating, there appear satellites at positions $h \pm \frac{1}{Q}$, $h \pm \frac{2}{Q}$, .. , where Q is the number of lines in the period of variation. The structure amplitudes of the main reflexions and the satellites are given by $J_0(hQb/a)$, $J_1(hQb/a)$, $J_2(hQb/a)$, ..., respectively, where $J_n(hQb/a)$ is the nth order Bessel function of the argument hQb/a, and b is the amplitude of the variation of the spacing, a.

If
$$\frac{hQb}{a} \ll 1, \tag{1}$$

one may put $J_0 = 1$, $J_1(hQb/a) = hQb/2a$ and $J_{n>1}(hQb/a) = 0$, that is, only the first pair of satellites is appreciable.

The intensities of the different order satellites will be given by J_0^2, J_1^2, J_2^2, ... It is interesting that the total intensity of a main reflexion and all its satellites is always equal to the intensity of the diffraction line from the undisturbed grating, because for any argument x

$$J_0^2(x) + 2\sum_{n=1}^{\infty} J_n^2(x) = 1 \tag{2}$$

(Whittaker & Watson 1927).

The side-bands on a powder photograph are made up of the satellites due to variation in the h, k and l directions. It has been found convenient to use as basis for the calculations I/I_0, the intensity of the first order side-band over that of the main line. According to the approximate treatment based on condition (1) this quantity should be

$$\frac{I}{I_0} = (h^2 + k^2 + l^2)\left(\frac{Qb}{2a}\right)^2 = N\left(\frac{Qb}{2a}\right)^2. \tag{3}$$

Equation (3) shows that the quantity I/I_0 increases rapidly with the order of reflexion. The rigorous treatment shows that I/I_0 should increase even more rapidly because equation (2) proves that the total intensity of main line plus side-bands is not altered by a sinusoidal variation. In consequence, if the side-bands are strong, the main line is weakened.

The results of the measurements seemed at first not to agree with the theory. It was found that I/NI_0 is not constant but decreases with N. Empirically the intensities can be rendered by an expression such as

$$\frac{I}{I_0 N} \propto e^{-\alpha N}, \tag{4}$$

where α is a constant.

Extinction

Extinction influences the ratio of the intensity of the side-bands to the main line because it weakens strong X-ray reflexions more than weaker ones. Therefore it makes weak lines or, in our case, side-bands appear relatively too strong compared with the main reflexions. In the present case this produces a systematic alteration of the relevant expression I/I_0N.

The presence of extinction in the specimens used was proved by intensity measurements on the same specimen cold-worked and annealed. It was assumed that cold working would remove, or at least reduce, extinction. A cold-worked specimen of Cu_4FeNi_3 was therefore photographed together with a platinum wire, annealed at 900° C, and again photographed together with the platinum wire, both being carefully replaced in the same position as the first time. The intensities of the diffraction lines from the platinum wire were used as standards. The specimen and platinum wire were mounted on the same specimen holder, the specimen being held in a hole through the centre of the holder, while the platinum wire was fixed on the holder as near as possible to the specimen. The two revolved round an axis between them, care being taken that they were both immersed in the X-ray beam at every position. Two specimens were used. one a solid rod cut out of a rolled sheet of alloy, the other filed powder, photographed through a silica tube of a wall thickness of about 0·02 mm.

This experiment showed that by the annealing the intensities of the X-ray reflexions were weakened. Low-order reflexions were reduced more than high ones. The lowest reflexion, 111, was reduced in one case to 0 6, in the other to 0·4 of its intensity in the cold-worked state.

These experiments indicate that there is extinction present In order therefore to derive numerical results from the experimental data it was necessary to work out an analytical treatment in order to correct for extinction. This treatment is based on a paper by Bragg & West (1929).

It is obvious that the extinction effect will decrease with N, since the higher orders are necessarily weaker than the lower ones, and so the higher orders give values of Qb/a which are more nearly correct than those given by the lower ones. If one could use really high values of N, the extinction would be zero, but since this is not possible, an attempt was made to find the true value of Qb/a by extrapolation of the observed results to $N = \infty$. For this purpose it is necessary to find a relation that gives a linear extrapolation against a function of N that is not infinite at $N = \infty$. In the treatment that follows, such a relation has been deduced

Analytical treatment of extinction

There are two kinds of extinction possible, primary and secondary. Primary extinction occurs in large perfect blocks, while secondary extinction (as well as primary extinction) may occur in mosaic crystals. Both kinds give a fairly similar variation with angle.

An approximate allowance for secondary extinction can be made by putting for the integrated reflexion under the influence of extinction

$$\rho' = \frac{\rho}{1+2g\rho}. \tag{5}$$

Here ρ is the true integrated reflexion and g is a constant characteristic of the extinction.

For the following the calculations have to be made in terms of the structure amplitude F. ρ is connected with F by the formula

$$\rho \propto \frac{\lambda^3}{\mu}\frac{1+\cos^2 2\theta}{\sin 2\theta} F^2, \tag{6}$$

where λ is the X-ray wave-length, μ the absorption coefficient and θ the Bragg angle. F', the value that follows for the structure factor if the measurements are not corrected for extinction, is connected with ρ' by the same formula. By substituting F and F' for ρ and ρ' in formula (5) we get

$$F'^2 = \frac{F^2}{1+\text{const.}\dfrac{\lambda^3}{\mu}\dfrac{1+\cos^2 2\theta}{\sin 2\theta} F^2}. \tag{7}$$

It can be seen that the larger F^2 is the more F'^2 is reduced However, the relationship is complicated by the factor $(1+\cos^2 2\theta)/\sin 2\theta$, which means that extinction will be relatively less for reflexions with θ near $\frac{1}{4}\pi$ than for reflexions with higher and lower θ. The factor λ^3/μ is practically constant for different radiations (Compton & Allison 1936) unless the value of μ is increased anomalously by the presence of an absorption edge. The anomalous absorption of iron for copper radiation has the consequence that λ^3/μ, and therefore extinction, is somewhat less for copper radiation than for the other radiations used. In spite of this it was found sufficient for an approximate treatment to set the factor $(\lambda^3/\mu)(1+\cos^2 2\theta)/\sin 2\theta$ equal to a constant. Thus

$$F'^2 = \frac{F^2}{1+cF^2}. \tag{8}$$

For the present problem the calculations are simplified by the fact that F, the structure amplitude not influenced by extinction, can be expressed as a simple function of the order of reflexion, N. As the alloy Cu_4FeNi_3 is face-centred cubic, F_0, the structure amplitude corrected for the temperature factor, is for all reflexions equal to $4f_0$, where f_0 is the mean atomic structure factor. f_0 itself is still a function of $\sin\theta/\lambda$ or N and falls off with increasing N. One can put with reasonably good approximation

$$F^2 = Ae^{-\epsilon N}, \tag{9}$$

where the constant ϵ includes the temperature factor and the variation of f_0 with N. Consequently

$$F'^2 = \frac{Ae^{-\epsilon N}}{1+cAe^{-\epsilon N}}$$

or
$$\frac{F'^2}{F^2} = \frac{1}{1+Ke^{-\epsilon N}}, \tag{10}$$

where K and ϵ are constants.

Equation (10) can be used to estimate the influence of extinction on I/I_0 and to correct for extinction. Extinction will weaken the intensities of the main lines I_0 according to equation (10) to I'_0 where

$$I'_0 = \frac{I_0}{1+Ke^{-\epsilon N}}. \tag{11}$$

The side-bands are considerably weaker than the main lines and are also broadened, so that they will be reduced much less. In the following it will be assumed that they are not affected by extinction. If condition (1) is fulfilled $I/I_0 = N(Qb/2a)^2$. Extinction will modify this to

$$\frac{I}{I'_0} = N\left(\frac{Qb}{2a}\right)^2(1+Ke^{-\epsilon N}). \tag{12}$$

This expression proves suitable to extrapolate to $N = \infty$. Plotting I/I'_0N against $e^{-\epsilon N}$ should give a straight line. The intercept for $e^{-\epsilon N} = 0$ ($N = \infty$) gives $(Qb/2a)^2$.

The value of the constant ϵ has to be calculated from a specimen without extinction. In the following the highest value measured on a cold-worked specimen will be used. An error in the assumption of ϵ fortunately does not affect the value of $(Qb/2a)^2$ extrapolated.

Comparison with experiment

Figure 1 shows extrapolation curves I/I'_0N as function of $e^{-\epsilon N}$ for some specimens. One can see that the linearity of the relation is reasonably well obeyed although one cannot claim much accuracy (as formula (10) is only roughly true). But this does not affect the calculation of Qb/a unduly because the method for correcting for extinction is an extrapolation to $N = \infty$. The intensities of reflexions {331} and {420} are most important for the extrapolation. It has been calculated by the most accurate formulae available (formula (7) and the table for primary extinction in *Internat. Tab.* 1935) that the intensities of these reflexions may be reduced with the strongest extinction observed by a factor less than 1·5. Altogether, taking into account all the errors of measurement, the error of a determination of Qb/a should be less than $\pm 25\,\%$.

The above observations show that extinction is an effect not to be neglected. It can affect the relative intensities of weak and strong reflexions. Besides, a closer investigation of formulae (10) and (12) shows that it can also alter the temperature factor and the broadening of lines measured. In the present case it completely falsified the evidence of the intensities, because they decrease systematically with angle.

Results

The correction for extinction makes it possible to give a quantitative description of the structure that produces the 'side-band' pattern. This structure is formed by a periodic variation of the lattice parameter in the direction of the cube edges. The period Q of the variation can be calculated from the position of peak intensities of the side-bands, while Qb/a follows from the intensities. The breadth of the side-bands shows that Q is not a constant but that a range of different values is present. If $(hQb/2a) < 1$ for all values of Qb/a present it can be shown that the statistical average

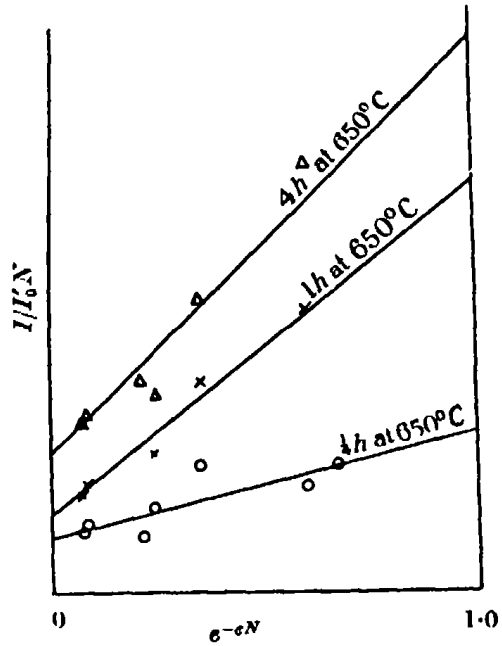


Figure 1. Some extrapolation curves for the calculation of Qb/a.

of the values of Q present can take the place of a constant period in all the calculations. It has to be noted that the position of the peak of the intensities of the side-bands, from which Q is calculated, does not necessarily give this average, as the higher values of Q give stronger side-bands and so are weighted more heavily; but the difference should not be large. In principle it would be possible to deduce from the broadening of the side-bands a distribution function of Q. In practice this cannot be done with any accuracy, but it has been estimated that the values of Q vary within about $\pm 50\,\%$ of the average value.

The periodic variation may be essentially one-dimensional so that lamellae normal to the three cube axes are present in each crystallite, or else it may be essentially three-dimensional, so that spherical aggregates of larger and smaller spacing alternate in a three-dimensional lattice. The difference between the two cases influences equation (3). The equation in the form in which it is written follows for a three-dimensional variation, while for the lamellar case I/I_0 is only $\frac{1}{3}N(Qb/2a)^2$.

It was thought interesting to follow the change of the periodic structure with time of annealing, as this provides rather detailed information about the changes of concentration in the alloy on a scale of 10^{-6} cm., which is otherwise not easily accessible. A series of photographs was therefore taken with different times of annealing at temperatures 550, 650 and 750° C. It was found that for a considerable time interval the side-band pattern remains, but that the separation of the side-bands decreases (figure 2 *a, b, c*). This stage could be analysed quantitatively. It corresponds to values of Qb/a for which condition (1) is valid. For longer times of annealing the pattern changes; higher order side-bands appear, which are in good agreement with the rigorous theory. Unfortunately, the diffraction lines become too complicated to be analysed quantitatively.

Table 1 shows the quantitative results of the measurements. The most striking point about these results is that b, the amplitude of the variation, remains constant within experimental error. The values of b found can be related to Δa, the difference of spacing of the two face-centred cubic phases in equilibrium. It seems reasonable to assume that the difference of volume between the two phases in the equilibrium state must be at least as large as that between a copper-rich and a copper-poor aggregate in the metastable periodic state. This gives the relation that for a lamellar variation $b \leqslant \frac{3\pi}{4}\Delta a$ while for a three-dimensional variation $b \leqslant \frac{\sqrt{3}\pi}{4}\Delta a$. It can be seen that for all three temperatures b is about equal to $\frac{\sqrt{3}\pi}{4}\Delta a$.

Table 1

t hr.	Q unit cells	b Angstrom units	Qb/a	Qb/a from second side-bands	
			550° C		
2½	28	—	—	—	Average $b = 0{\cdot}049$ A
5	37	0·052	0·54	—	$\frac{\sqrt{3}\pi}{4}\Delta a = 0{\cdot}046$ A
10	38	0·047	0·50	—	
24	52	0·053	0·77	—	In two-phase state
56	59	0·042	0·69	0·6–0·8	$\Delta a = 0{\cdot}033_7$ A
120	—	—	—	0·9	
			650° C		
¼	35·5	0·0395	0·39	—	Average $b = 0{\cdot}035$ A
½	43	0·038	0·46	—	$\frac{\sqrt{3}\pi}{4}\Delta a = 0{\cdot}035$ A
1	50·5	0·0325	0·465	—	
2	56	0·034	0·53	—	In two-phase state
4	69	0·0325	0·63	—	$\Delta a = 0{\cdot}025_7$ A
8	79	0·033	0·73	0·6–0·75	
16	—	—	—	0·8–0·9	
			750° C		
¼	65	0·017	0·31	—	Average $b = 0{\cdot}018$ A
½	73	0·019	0·38	—	$\frac{\sqrt{3}\pi}{4}\Delta a = 0{\cdot}018$ A
1	81	0·017	0·40	—	
5	111	—	—	0·6–0·7	In two-phase state
					$\Delta a = 0{\cdot}013_4$ A

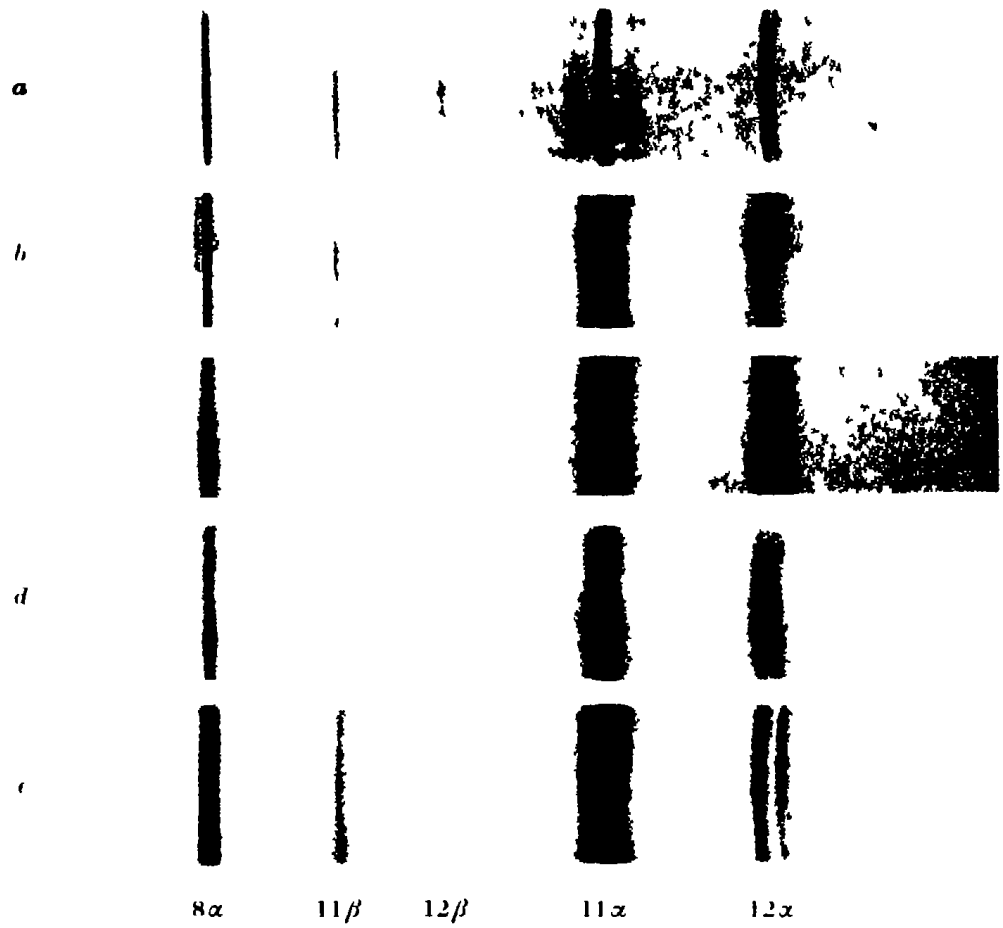


FIGURE 2. X-ray powder photographs (Fe Kα radiation) of specimens which have been annealed at 950° C, quenched and annealed again at 650° C for (*a*) ½ hr, (*b*) 1 hr, (*c*) 8 hr, (*d*) 16 hr, (*e*) 280 hr.

The average period Q increases with time, and figure 3 shows that there is a linear relation with the logarithm of time at all three temperatures observed. It is interesting that the slope of the curves of Q as function of $\log t$ seems to be independent of temperature. The evidence of figure 3 can thus be represented by the equation

$$Q = A \log t + \phi(T), \tag{13}$$

where t is the time, A a constant and $\phi(T)$ some function of the temperature.

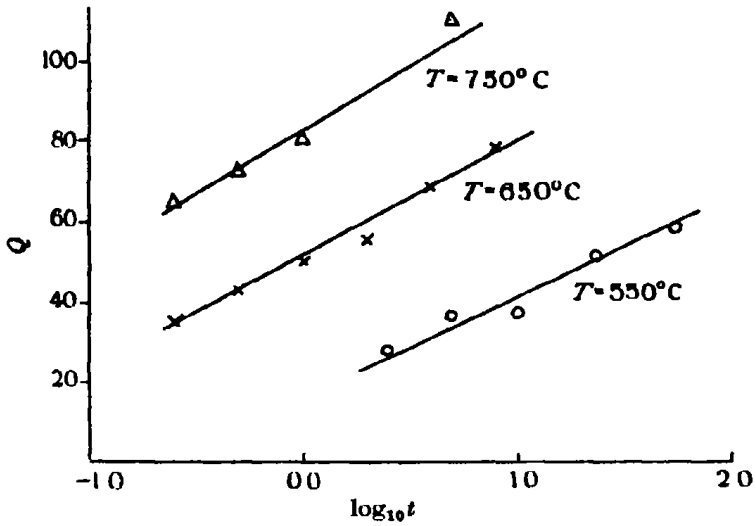


FIGURE 3. The wave-length Q of the periodic variation as function of the temperature and the logarithm of the time of annealing.

Column 4 of the table shows Qb/a as deduced from the intensities. The calculations are made with the assumption of a lamellar variation, that is, from $\frac{I}{I_0} = \frac{N}{3}\left(\frac{Qb}{2a}\right)^2$. If a three-dimensional variation was assumed, Qb/a would appear smaller by a factor of $\sqrt{3}$. This factor provides a possibility of distinguishing between the two cases, as, for some specimens with Qb/a on the limit of validity of condition (1), it is possible to estimate Qb/a from the appearance of secondary side-bands. These estimated values are given in column 5. They make a lamellar variation probable, but the evidence is not quite sufficient to settle the point, particularly as the evidence of b as compared with Δa is indecisive.

The quantitative results given in the table are reproducible within experimental error. Specimens of solid metal and rods of sintered powder give the same results.

After longer times of annealing secondary side-bands appear and the side-band pattern gradually changes into another diffraction pattern (see figure 2 *d*, *e*, plate 12). This pattern (figure 2 *e*) has been explained by Bradley (1940) as being due to tetragonal lamellae. A consideration of the higher order Bessel functions shows that it may also be due to a periodic variation with a range of different values of Q, when Qb/a is large.

If Qb/a is large, each reflexion will have different order satellites, with structure amplitudes given by $J_n(hQb/a)$. We shall put $(hQb/a) = x$. As it has been found that b is constant, and h is also constant for any particular reflexion, x is proportional to Q. The position of each satellite is given by $h \pm n/Q$. This means that the separation of a satellite from the main line will be proportional to n/x. In order to follow the amplitude as function of the separation of satellites $J_n(x)$ has been plotted against n/x for different constant values of x (figure 4).

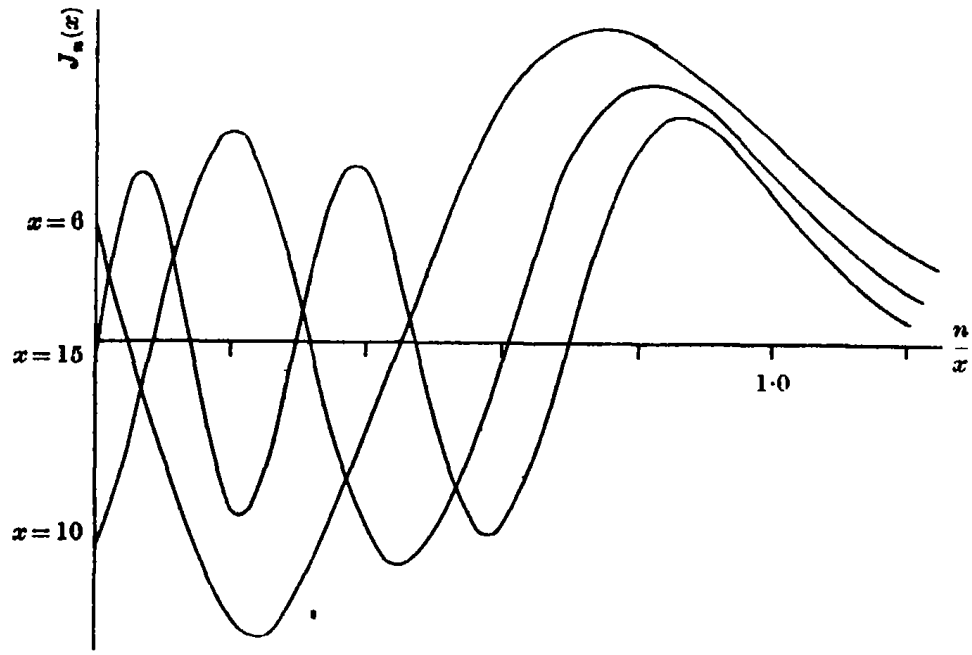


FIGURE 4. Bessel functions $J_n(x)$ as function of n/x for $x=6$, $x=10$, and $x=15$.

If the structure amplitude of the whole structure is made up from the amplitudes given by the different values of Q present, it can be seen that for $(n/x) < 1$ these amplitudes will in general cancel out except for values near $(n/x) = 1$. This means that in general the intensity of a reflexion will be appreciable only at a position corresponding to $(n/x) = 1$, that is at a distance of $\pm\left(\frac{hQb}{a}\Big/Q\right)$ from the position of the reflexion from the undisturbed lattice. In consequence in the one-dimensional case each reciprocal-lattice spot of the undisturbed lattice will be replaced by two diffuse spots displaced from the original spot by approximately hb/a along the axis. In three dimensions this is very nearly the same reciprocal lattice as that given by two tetragonal phases fitting together on a (100) plane, as postulated by Bradley.

KINETIC CONSIDERATIONS

The periodic variation of the lattice parameter that has been found must be due to differences of concentration set up at regular intervals. The effect probably has

some analogy with the Liesegang phenomenon. It is interesting that it occurs reproducibly in all specimens. The reproducibility of the phenomenon is perhaps connected with the great perfection of the crystallites in the alloy, which is indicated by the presence of extinction. The dissociation occurs in perfect crystallites of a length of the order of 10^{-4} cm., a size which is large compared with the scale on which the dissociation occurs. Thus the influence of mosaic faults and mechanical deformation is presumably eliminated.

The fact that b is constant shows that the maximum difference of concentration is reached even with the shortest times of annealing given. This means that the changing of the X-ray pattern with the time of annealing is not due to the initial setting up of concentration differences but corresponds rather to the phenomenon of grain growth. The value of Q, i.e. the grain size, is determined by time and temperature of annealing.

It is interesting that equation (13) separates the variables time and temperature It would be interesting to know whether this represents a general relationship in the phenomenon of grain growth.

The equation (13) obviously cannot hold for the whole of the process of dissociation, as it gives $Q = -\infty$ for $t = 0$ Some other relation must therefore hold for the initial setting up of concentration differences. Formula (13) can also not hold for longer times of annealing, because, if one extrapolates Q as function of $\log t$ one sees that $Q = 1000$ should be reached at 650° C only in 10^{30} years, while a sharp 2-phase pattern, indicating a grain size of about 10^{-4} cm , is already reached in 11 weeks. The extraordinary slowness of the growth of Q is probably due to the geometrical arrangement of the aggregates, which are 'interlocked', and not to abnormally slow diffusion. This is suggested also by an experiment which was made in order to compare the rate of forming and destroying the periodic state. An alloy near the composition of Cu_4FeNi_3 which is single phase at 750° C, was brought into the periodic state by annealing for 6 days at 550° C. By annealing at 750° C it could be made single phase in less than 5 min., although Cu_4FeNi_3, which is in equilibrium 2 phase at 750°, is still in the periodic state even after several hours at 750° C.

Our thanks are due to Professor Sir Lawrence Bragg, F.R.S , and to Dr A. J. Bradley, F.R.S , for their encouraging interest in this work. Financial assistance has been received from the British Electrical and Allied Industries Research Association and from the Iron and Steel Federation, and one of us (V. D.) wishes to thank Girton College for a Research Studentship.

References

Bradley, A. J. 1940 *Proc. Phys. Soc.* **52**, 80.
Bragg, W. L. & West, J. 1929 *Z. Krist.* **69**, 118.
Compton, A. H. & Allison, S. K. 1936 *X-rays in theory and experiment*, p. 534.
Daniel, V. & Lipson, H. 1943 *Proc. Roy. Soc.* A, **181**, 368.
International tables for the determination of crystal structures 1935.
Whittaker & Watson 1927 *Modern analyses*, p. 379.

Dielectric properties of dipolar substances†

By H. Frohlich and R. Sack

H. H. Wills Physical Laboratory, University of Bristol

(*Communicated by N. F. Mott, F.R.S.—Received* 30 *July* 1943)

From an investigation of structure it is shown that there exists a large group of dipolar organic solids whose dipoles have two equilibrium positions with opposite dipole direction. To calculate the dielectric properties Onsager's theory has been extended and developed into a systematic approximation which converges above a critical temperature. To derive the local field acting on a dipole we have replaced the surroundings by a continuum whose dynamic dielectric properties we have taken into account As a result we find larger dielectric constants and smaller dielectric losses than in Onsager's theory. We have also shown that liquids with high viscosity behave similarly to solids, while for liquids with low viscosity there are no such deviations from Onsager's theory.

I. Introduction

1. The purpose of this paper is twofold First, from an analysis of their structure we shall show that there exists a large group of dipolar solids whose dipoles do not rotate but have two stable positions with opposite dipolar direction. Secondly, we shall develop into a systematic theory a method which has been widely used in order to account for dipolar interaction. It is the method whereby the local field acting upon a dipole is obtained by replacing by a continuous medium the surroundings of this dipole outside a cavity. Debye (1929), who first applied this method to dipolar substances, assumed that the local field is given by the usual Lorentz theory. His local field is thus deduced on the assumption that the dielectric polarization inside the cavity is identical with the average macroscopic polarization. Onsager (1936) pointed out that this was incorrect when a permanent dipole is being placed inside the cavity. He showed that the local field is composed of two components: (i) the cavity field which is the field produced inside the cavity by the external field on the assumption that the dipole has been removed; (ii) the reaction field which is the field at the position of the dipole which is produced through the action of the dipole on its surroundings. Onsager then assumed that this reaction field has always the same direction as the dipole. Thus it does not exert any force upon the dipole and can be omitted. This procedure is open to objections (see also Sauer & Temperley 1940). It is true the equilibrium value of the reaction field has the direction of the dipole. This equilibrium value will, however, never be established in view of the frequent changes of the direction of the dipole. It will thus be necessary to take the reaction field into account, and to calculate it from the dynamic dielectric properties of the surrounding medium. Nevertheless, we shall find that Onsager's theory constitutes a correct zero approximation although for reasons different from

† Parts of this paper are based on reports L/T. 124, L/T. 132 and L/T. 142 (by H. F.) of the British Electrical and Allied Industries Research Association (E.R.A.).

those given by him. We shall see that, for the substances which we treat here, the method of replacing the surroundings of a dipole by a continuous medium is restricted to temperatures above a critical temperature T_0, where the energy kT_0 is of the order of magnitude of the dipolar interaction. Thus in zero order this interaction, and hence the reaction field, can be neglected. Clearly the deviations of the more exact theory from Onsager's theory increase with decreasing temperature. While they may be appreciable for solids and for liquids with high viscosity we shall find that for liquids with low viscosity Onsager's theory is always a very good approximation.

In the following we shall start with a discussion of the structure of substances consisting of long-chain molecules. We shall then develop the general mathematical method and finally use it to calculate the dielectric constant and the dielectric loss.

II. Model and structure

2. *Solids.* The structure of many of the substances with which we are concerned in this paper can be derived from the structure of paraffins which has been investigated by Müller (1928). Paraffins are long-chain molecules forming a plane zigzag with a CH_2 group at each corner and a CH_3 group at the ends. Polar molecules can be obtained from paraffins by introducing polar groups. Thus a ketone is obtained by replacing a CH_2 group by a C=O group (figure 1). In the paraffin structure the chains are arranged in layers whose thickness is approximately equal to the chain length. Within such a layer the molecules form rectangular cells with side lengths a, b, c where $a \sim 5$ A, $b \sim 7{\cdot}5$ A and c slightly larger than the chain length. Figure 2 shows how the chains, whose cross-sections are indicated, intersect the a-b plane. The next layer on top of the one considered just now is slightly shifted in the b-direction From this structure other long-chain structures can be obtained by choosing as unit cell a parallelepiped with arbitrary angles.

All these lattice structures strongly suggest that each chain has a second position of equilibrium obtained by turning the chain plane by 180°, thus reversing the direction of the dipole† (see figure 2). The existence of two positions for each molecule should lead to order-disorder transitions connected with an anomaly in the specific heat as has actually been observed for paraffins (see Muller 1932; Ubbelohde 1938). The transition from one state of a molecule into the other has been investigated previously (Frohlich 1942) for the case of dipolar chain molecules dissolved in paraffin wax.

Let us now restrict ourselves to molecules which contain only a single dipolar group which usually lies in the chain plane In this case the position of a molecule can be described in a unique way by the position and direction of a single dipole. We may thus speak of dipolar interaction when we think of that part of the interaction between molecules which depends on the direction of the chain planes.

† This would probably entail some displacement of the molecules.

The conception of two stable dipolar directions has already been used by Debye (1929), but it has not been brought into connexion with any actual structure. Usually the conception of rotating dipoles (cf. Pauling 1930) is being used. While we do not doubt that such molecules as HCl have to be considered as rotating dipoles,

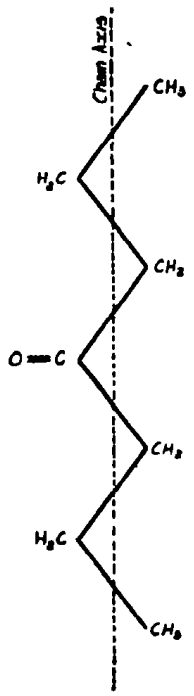


FIGURE 1. Model of chain with a ketone group.

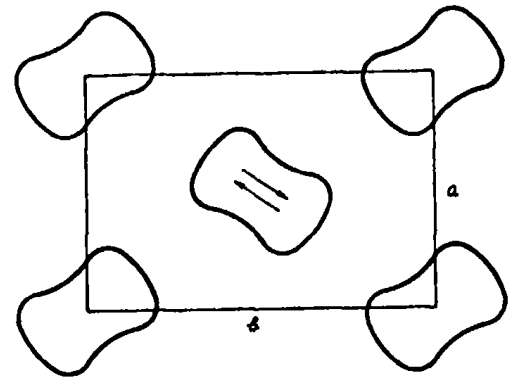


FIGURE 2. Plan of arrangement of chains in a plane perpendicular to the chain axes The two possible directions of a dipole are indicated for the central molecule.

even in the solid state there is certainly a large group of solids, the dipoles of which do not rotate. For these substances we shall adopt the following model:

(i) The substance consists of a continuous medium containing dipoles which form some crystalline lattice. The continuous medium has the real dielectric constant ϵ_∞ independent of the frequency of an external field, i e. the corresponding polarization is loss-free and relatively inertia-free.

(ii) All dipoles have the dipolar momentum μ. Apart from their mutual interaction they are subject to non-dipolar forces which tend to orient them into either of two opposite directions. We shall call them stable positions or equilibrium positions of the dipoles. The potential energy of the dipoles is supposed to be the same for the two positions if the mutual interaction between dipoles is being neglected.

(iii) Occasionally a dipole will have sufficient kinetic energy to jump over the potential hill separating the two equilibrium positions. Let p be the probability per second for such a transition, neglecting dipolar interaction. We then define a time of relaxation τ of the dipole by

$$2\tau = 1/p. \qquad (2{\cdot}1)$$

(iv) In an external electric field E a dielectric polarization M is induced per unit volume defining the dielectric constant ϵ by

$$M = \frac{\epsilon - 1}{4\pi} E. \tag{2·2}$$

If P denotes the contribution of the dipoles to M then

$$M = P + \frac{\epsilon_\infty - 1}{4\pi} E. \tag{2·3}$$

In alternating fields it is useful to introduce complex quantities. Let ω be the angular frequency of the field, then we shall write

$$E = E_0 e^{i\omega t} \tag{2·4}$$

and

$$\epsilon = \epsilon' - i\epsilon'', \tag{2·5}$$

where ϵ' and ϵ'' as well as ϵ_∞ are real. This way of writing means that the field strength is the real part of (2·4) while the electric displacement is the real part of ϵE. Hence the dielectric loss per second and per unit volume is given by

$$\tfrac{1}{2}\epsilon'\omega E_0^2 \tan\theta$$

where the loss angle θ is given by $\tan\theta = \dfrac{\epsilon''}{\epsilon'}$. (2·6)

Our way to introduce the non-dipolar contributions to the polarization (i.e. $(\epsilon_\infty - 1)E/4\pi$) differs from the usual way where a molecular polarization is first introduced and connected with ϵ_∞ with the help of the Clausius-Mosotti formula (see Debye 1929). The use of this formula is, however, objectionable (see, for instance, Kurtz & Ward 1937).

3. *Liquids.* From X-ray investigations it is known that in a liquid the neighbours of a given molecule are mostly arranged according to a certain crystalline structure except for some small deviations. Accordingly we shall assume that, even in a liquid, a molecule, and hence a dipole, has two stable positions relative to its neighbours. Thus if we approximate the surroundings of a molecule by a macroscopic liquid this must be assumed to adhere to the surface of our molecule. Thus, if we neglect the transitions of the dipole between its two equilibrium positions, we are led to Debye's model where the motion of a dipolar molecule in a liquid is treated in the same way as the motion of a macroscopic body (Debye 1929). Let τ_D be the time of relaxation according to Debye which is proportional to the viscosity of the liquid. Then whenever $\tau_D \ll 1/p$ (see equation (2·1)), as is probably the case for liquids with small viscosities, Debye's model is applicable. For substances with very high viscosity, however, probably $\tau_D \gg 1/p$. In this case, therefore, our solid model is more appropriate. For not too high temperatures, this should apply in particular to amorphous organic substances which have no sharp melting-point.

III. The mathematical method

4. It is the purpose of this section to calculate the dipolar contribution to the dielectric polarization induced by the external field E. For solids we shall always assume that we are dealing with polycrystalline material and, therefore, average over the angle θ between field and dipolar direction. If we neglect the interaction between dipoles, the energy of one dipole becomes $\mu E \cos\theta$. For weak static fields ($\mu E \ll kT$) the Boltzmann theorem yields at once for the polarization per unit volume

$$\mathbf{P} = \frac{\mu^2 n \mathbf{E}}{3kT}, \tag{3·1}$$

where n is the number of dipoles per unit volume. To neglect the dipolar interaction is, however, permissible only in the relatively uninteresting case $\epsilon - \epsilon_\infty \ll 1$. The main difficulty to account for this interaction is its long range $\sim 1/r^3$ which makes it impossible to take into consideration the interaction between nearest neighbours only.

In order to simplify this problem we shall follow a method in which the action upon a selected dipole of the external field and of all dipoles is being described by a local field **F** which has been calculated in a simplified manner It is then assumed that on an average all dipoles behave in the same way as the selected one The polarization is thus obtained from (3·1) if **E** is replaced by **F**. To determine **F** we surround the selected dipole by a cavity of volume $1/n$ and treat the outside as a continuous medium whose dielectric properties are the same as the macroscopic dielectric properties which we want to calculate Following Onsager (1936) we split **F** into two parts: (i) the cavity field **G** which is obtained on the assumption that the dipole has been removed from the cavity; (ii) the reaction field **R** which is the change of the field at the position of the dipole through its action upon the surrounding dipoles,

$$\mathbf{F} = \mathbf{G} + \mathbf{R}. \tag{3·2}$$

Our definition of G is slightly different from Onsager's who removes not only the dipole from the cavity in order to determine G but the whole molecule. Thus the dielectric constant of the empty cavity is ϵ_∞ in our case, but unity in his. The actual value of G depends on the shape of the cavity. For a sphere of radius a, given by

$$\frac{4\pi}{3} a^3 = \frac{1}{n}, \tag{3·3}$$

we obtain, following Onsager's calculation,

$$\mathbf{G} = \frac{3\epsilon}{2\epsilon + \epsilon_\infty} \mathbf{E}, \tag{3·4}$$

independent of a. It should be noticed that this expression remains correct for alternating fields if ϵ represents the complex dielectric constant at the frequency in question.

Our treatment of the reaction field R is essentially different from Onsager's who assumes that R is equivalent to R_1, the value which R tends to reach if the dipole remains in a given direction for a time which is long compared to the relaxation time τ. This, however, is not the case, and, therefore, we shall describe R by its dynamic properties, i.e. by its time-dependent differential equation. From the definition of the time of relaxation it follows that

$$\tau \frac{d\mathbf{R}}{dt} = -\mathbf{R} + R_1 \frac{\boldsymbol{\mu}}{\mu}, \tag{3·5}$$

an equation which we shall justify below (see text between equations (3·17) and (3·18)). $\boldsymbol{\mu}/\mu$ is the unit vector in the direction of the dipole (which depends on time) and obviously

$$|\mathbf{R}| < R_1, \tag{3·6}$$

unless the dipole remains infinitely long in the same direction. For a spherical cavity R_1 can be calculated similarly to Onsager's calculation of his R Given a sphere of radius a with the static dielectric constants ϵ_∞ and ϵ_s inside and outside the sphere (ϵ_s is the static value of ϵ) Let ψ be the electric potential so that

$$\psi = \mu \cos\theta / \epsilon_\infty r^2$$

would be a solution if $\epsilon_\infty = \epsilon_s$, i.e if there were no dipoles present (r = distance from the centre of the sphere, θ = angle between $\mathbf{r}$ and $\boldsymbol{\mu}$.) If, however, $\epsilon_s \neq \epsilon_\infty$ we find from $\nabla^2\psi = 0$,

$$\psi = \left(\frac{\mu}{\epsilon_\infty r^2} - R_1 r\right)\cos\theta, \quad \text{if } r < a,$$

$$\psi = \frac{\mu^*}{\epsilon_s r^2}\cos\theta, \qquad \text{if } r > a$$

The constants R_1 and μ^* have to be determined from the boundary conditions at the surface of the sphere (continuity of the tangential component of the electric field strength and of the normal component of the electric displacement). This leads to

$$R_1 = \frac{2(\epsilon_s - \epsilon_\infty)}{\epsilon_\infty(2\epsilon_s + \epsilon_\infty)}\frac{\mu}{a^3}. \tag{3·7}$$

Onsager treats R not as a dynamic but as a static quantity, i.e he uses $\mathbf{R} = R_1\boldsymbol{\mu}/\mu$. The energy $-(\boldsymbol{\mu}, \mathbf{R})$ of the dipole in the field $\mathbf{R}$ is in his case $-\mu R_1$. It is thus independent of the dipole direction and can be omitted. This would be correct only if $1/\tau$ were large compared to $d(\boldsymbol{\mu}/\mu)/dt$. This is the case only for the non-dipolar part of Onsager's reaction field. In our treatment of the reaction field we have defined R as being due to dipoles only, while the non-dipolar contributions, described by ϵ_∞, have been assumed to be free of inertia.

5. *Range of validity.* Before starting with the actual calculation we want to investigate the range of validity of our method. One of the main points of the method is to calculate the reaction field with the help of the macroscopic dielectric constant

of the substance making use of the fact that in the absence of an external field the polarization vanishes everywhere if we imagine the dipole to be removed from the cavity. In the presence of the dipole an electric field is set up whose order of magnitude is R in the neighbourhood of the cavity. Now consider a small region of the continuum in the neighbourhood of the cavity co-ordinated to a selected dipole. This region is being polarized through the action of this dipole. We might, however, have selected any other dipole which would induce a different polarization in this region. Our method will work only if these various polarizations do not influence each other, i.e. if they superpose linearly. This is the case if we do not approach saturation Since saturation is approached if the local field is larger than kT/μ, our condition is $\mu R < kT$ or with (3·6), and introducing a temperature $T_0 = \mu R_1/k$,

$$\mu R_1 = kT_0 < kT. \tag{3·8}$$

A similar condition is obtained from a consideration of the cavity field G. From (3·4) we see that G vanishes in the absence of an external field. Thus the field at the position of a dipole should be due to the reaction field only, i.e. it should vanish if we imagine this dipole to be removed. In the average this should be the case for liquids and for the disordered state of the dipoles in solids, but in general not for the ordered state. Thus if the order-disorder transition† occurs at the temperature T_1, we require $T > T_1$. kT_1 is of the order of magnitude of the energy required to turn at $T = 0$ a dipole from one equilibrium position into the other. If we assume that this energy is due to the electric interaction between the dipoles then $T_1 \sim T_0$.

Equation (3·8) has been derived as a necessary condition for the application of our method. It should be realized that we have no means of estimating the deviation of our approximation from the exact solutions. In connexion with this it is of importance that Van Vleck (1937*a*, *b*) has shown that the exact partition function of an assembly of dipoles can be developed into a convergent power series in $1/T$ if kT is large enough. For a model which corresponds to our low viscosity liquid model he found that his results are identical with Onsager's up to the third order in $1/T$. In § 7 we shall also find that Onsager's approximation is very good for such liquids but we shall find stronger deviations for solids. Van Vleck's method naturally is much more difficult than Onsager's, and it would probably be very complicated if applied to alternating electric fields. His method is of importance, however, as it provides a justification for the applicability of methods in which the surroundings of a dipole are replaced by a continuous medium (as used by Debye, Onsager and in the present paper).

6. *The fundamental equations. Solids.* We shall now derive the fundamental differential equations of our theory for the model of a solid dielectric, described in § 2, in which each dipole has two stable positions with opposite dipole direction. These we shall call the 1- and the 2-directions. Our aim is to calculate the polarization

† A study of order-disorder transitions in solid dielectrics of a type which we do not consider here was given by R. H. Fowler (1935).

P induced by the field E. If n_1 and n_2 are the numbers of dipoles per unit volume in the 1- and 2-directions respectively, clearly

$$P = \mu(n_1 - n_2)\cos\theta, \tag{3·9}$$

where θ is the angle between the field and the 1-direction over which we shall average (polycrystalline material). Throughout we shall assume weak external fields, i.e.

$$\frac{\mu E}{kT} \ll 1, \quad \text{and hence} \quad \frac{\mu G}{kT} \ll 1. \tag{3·10}$$

We also remember that according to (3·8) we can introduce a parameter $\gamma < 1$ by

$$\gamma = \frac{\mu R_1}{kT} < 1. \tag{3·11}$$

This enables us to build up a systematic approximation by developing all expressions into power series in γ. For the present we do not require the actual value of γ but only the fact that $\gamma < 1$, and we shall calculate the polarization P up to the first order in γ. In our zero order, γ, and hence the reaction field, is neglected entirely as in Onsager's theory. This theory is, therefore, a correct zero order approximation. The reason for this is, however, not the fact that the reaction field immediately follows the dipole and thus does not exert any force (Onsager), but it is due to $\mu R_1 < kT$ whenever our method is applicable.

Let us follow Debye (1929) in giving a brief derivation of the zero-order equations. If w_{12} and w_{21} are the probabilities per second for the dipolar transitions $1 \to 2$ and $2 \to 1$ respectively, clearly

$$\frac{dn_1}{dt} = -\frac{dn_2}{dt} = -w_{12}n_1 + w_{21}n_2, \tag{3·12}$$

and

$$n_1 + n_2 = n. \tag{3·13}$$

Now since we neglect R in zero order, the local field F is given by G, and the energy of a dipole is $\mp \mu G\cos\theta$ for the 1- and 2-directions respectively. In the equilibrium state and for static fields, $dn_1/dt = 0$ and thus $n_1/n_2 = w_{21}/w_{12}$. From Boltzmann's theorem we find

$$\frac{n_1}{n_2} = \frac{w_{21}}{w_{12}} = e^{2\mu G\cos\theta/kT},$$

or, using (3·10), and the fact that $w_{21}(G)$ must be equal to $w_{12}(-G)$,

$$w_{21} = ce^{\mu G\cos\theta/kT} \sim c\left(1 + \frac{\mu G\cos\theta}{kT}\right), \quad w_{12} = ce^{-\mu G\cos\theta/kT} \sim c\left(1 - \frac{\mu G\cos\theta}{kT}\right), \tag{3·14}$$

where c may depend on T. Assume now (3·14) to hold also for time-dependent fields G which is correct if the frequency of the field is small compared to the proper

frequencies of the dipole. Then inserting (3·14) into (3·12) we find, using (3·13) and (3·9), and neglecting quadratic terms in G,

$$\frac{dP}{dt} = -\frac{P}{\tau} + \frac{P_0}{\tau}, \tag{3·15}$$

where

$$\frac{1}{\tau} = 2c, \tag{3·16}$$

and, averaging over θ,

$$P_0 = \frac{\mu^2 nG\cos^2\theta}{k\bar{T}} = \frac{\mu^2 nG}{3k\bar{T}}. \tag{3·17}$$

Let us now turn to the first-order approximation taking into account the reaction field R. Since R itself is small of the first order it will be sufficient to calculate R from the zero-order equation. This we have already done in equation (3·5) which is identical with (3·15). Since in solids the dipole can have only two directions, (3·5) and (3·6) become

$$\left.\begin{aligned} \tau\frac{dR}{dt} &= -R + R_1, \quad \text{dipole in 1-direction} \\ \tau\frac{dR}{dt} &= -R - R_1, \quad \text{dipole in 2-direction} \end{aligned}\right\} \tag{3·18}$$

and

$$-R_1 < R < R_1. \tag{3·19}$$

In contrast to the zero-order approximation the energy of a dipole is now

$$\mp\mu(G\cos\theta + R),$$

where the $-$ stands for the 1-direction and the $+$ for the 2-direction. Consequently instead of (3·14) we have now, using (3·16),

$$w_{21} = \frac{1}{2\tau}e^{\mu(G\cos\theta + R)/kT}, \quad w_{12} = \frac{1}{2\tau}e^{-\mu(G\cos\theta + R)/kT},$$

from which we obtain approximately

$$w^+ = \tau(w_{21} + w_{12}) \sim 1 + \gamma\beta x, \tag{3·20}$$

$$w^- = \tau(w_{21} - w_{12}) \sim \beta + \gamma x,$$

where† $x = R/R_1$, and

$$\beta = \frac{\mu G\cos\theta}{k\bar{T}}. \tag{3·21}$$

In equation (3·12), which is an exact equation, the transition probabilities w_{12}, w_{21} now contain R given by the differential relations (3·18) which implicitly depend on n_1 and n_2. Thus in first order (3·12) cannot be solved so easily as in zero order. To find a solution we introduce a probability function $p_1(R)dR$ denoting the probability that at the time t a dipole is in the 1-direction while the reaction field has

† Our final result (i.e. the expression for P) will be proportional to β. We have thus to keep terms $\sim\beta\gamma$ but neglect terms $\sim\beta^2$ and $\sim\gamma^2$ in order to find P correct in the first order in γ.

a value between R and $R+dR$. Similarly we introduce $p_2(R)$ for the 2-direction. Thus

$$n_i = n\int_{-R_1}^{R_1} p_i(R)\,dR \quad (i = 1, 2). \tag{3·22}$$

p_i, and hence n_i, may depend on the time t. We shall now develop the kinetic equations for $p_i(R)$. The differential coefficient of p_i with respect to the time is composed of two terms

$$\frac{\partial p_i}{\partial t} = \left(\frac{\partial p_i}{\partial t}\right)_{\text{dipole}} + \left(\frac{\partial p_i}{\partial t}\right)_R. \tag{3·23}$$

The first term refers to a change in dipole direction keeping R constant, whereas in the second term R varies but the dipole direction is kept constant. Clearly in analogy to (3·12)

$$\left(\frac{\partial p_1(R)}{\partial t}\right)_{\text{dipole}} = -\left(\frac{\partial p_2(R)}{\partial t}\right)_{\text{dipole}} = -w_{12}p_1(R) + w_{21}p_2(R). \tag{3·24}$$

To find the second term we notice that in the time interval dt, R increases by $\dot{R}dt$ ($\dot{R} = dR/dt$). Thus $p(R)\,dR$ at the time t is equal to $p(R-\dot{R}dt)\,d(R-\dot{R}dt)$ at the time $t-dt$, i e developing $p(R)$,

$$p(R)\,dR\,|_t = p(R-\dot{R}dt)\,d(R-\dot{R}dt)\,|_{t-dt} = \left[p(R) - \frac{\partial p(R)}{\partial R}\,\dot{R}dt\right][dR - d\dot{R}dt]\,|_{t-dt}.$$

Now, neglecting dt^2 we obtain

$$\left(\frac{\partial p_i}{\partial t}\right)_R = [p_i(R)\,dR\,|_t - p_i(R)\,dR\,|_{t-dt}]\frac{1}{dR}\frac{1}{dt} = -\frac{\partial p_i}{\partial R}\dot{R} - p_i\frac{d\dot{R}}{dR}.$$

We now make use of equations (3·18) for $\dot{R}$ and find

$$\left(\frac{\partial p_1}{\partial t}\right)_R = \frac{\partial p_1}{\partial R}\frac{R-R_1}{\tau} + \frac{p_1}{\tau}, \quad \left(\frac{\partial p_2}{\partial t}\right)_R = \frac{\partial p_2}{\partial R}\frac{R+R_1}{\tau} + \frac{p_2}{\tau}. \tag{3·25}$$

Inserting (3·24) and (3·25) into (3·23) we obtain

$$\left.\begin{aligned}\tau\frac{\partial p_1}{\partial t} &= (R-R_1)\frac{\partial p_1}{\partial R} + p_1 + \tau w_{21}p_2 - \tau w_{12}p_1,\\ \tau\frac{\partial p_2}{\partial t} &= (R+R_1)\frac{\partial p_2}{\partial R} + p_2 - \tau w_{21}p_2 + \tau w_{12}p_1.\end{aligned}\right\} \tag{3·26}$$

To bring these equations into a mathematically more useful form we use the dimensionless variable x defined by

$$x = \frac{R}{R_1} \quad (-1 < x < 1), \tag{3·27}$$

and instead of p_1 and p_2 we introduce the functions

$$f(x) = (p_1 + p_2)R_1, \tag{3·28}$$

and

$$g(x) = (p_1 - p_2)R_1. \tag{3·29}$$

We then obtain from (3·26), using (3·20),

$$\tau\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}(xf-g), \tag{3·30}$$

$$\tau\frac{\partial g}{\partial t} = \frac{\partial}{\partial x}(xg-f) - w^{+}g + w^{-}f. \tag{3·31}$$

These two differential equations have to be solved under the following conditions. First, since the total number of dipoles per unit volume, $n = n_1 + n_2$, is a constant, we find from (3·28), (3·22) and (3·27)

$$\int_{-1}^{1} f(x)\,dx = 1. \tag{3·32}$$

Furthermore, in the expressions $(\partial p_i/\partial t)_R$ the dipolar direction is kept constant.

Therefore $$\int_{-R_1}^{R_1}\left(\frac{\partial p_i}{\partial t}\right)_R dR = 0 \quad (i = 1, 2).$$

Using equations (3·27), (3·28), (3·29), (3·25) as well as (3·32) this condition becomes

$$\int_{-1}^{1}\frac{\partial}{\partial x}(xg-f)\,dx = 0. \tag{3·33}$$

From the four equations (3·30)–(3·33) we have to determine the polarization P. Thus according to (3·9) we require $n_1 - n_2$ which, using (3·22), (3·27) and (3·29), is given by

$$n_1 - n_2 = n\int_{-1}^{1} g(x)\,dx = nJ. \tag{3·34}$$

To find the integral J we integrate equation (3·31) from $x = -1$ to $x = 1$. Inserting $w^{\pm}$ from (3·20) we obtain, making use of the conditions (3·32) and (3·33),

$$\tau\frac{dJ}{dt} = -J - \gamma\beta\int_{-1}^{1} xg\,dx + \beta + \gamma\int_{-1}^{1} xf\,dx.$$

Here the integral $\int xg$ has a factor $\gamma\beta$. The function $g(x)$ is, therefore, required in zero order in γ and β only, i.e. neglecting the effect of an external field as well as of the reaction field. In this approximation both $\partial f/\partial t$ and $\partial g/\partial t$ vanish if we assume equilibrium, and $w^{+} = 1$, $w^{-} = 0$. Equations (3·30) and (3·31) thus become

$$0 = \frac{\partial}{\partial x}(xf-g) \quad \text{and} \quad 0 = \frac{\partial}{\partial x}(xg-f) - g, \quad \text{if} \quad \gamma = 0,\ \beta = 0.$$

Multiplying the second of these equations by x and integrating yields

$$\int_{-1}^{1} xg\,dx = \int_{-1}^{1} x\frac{\partial}{\partial x}(xg-f)\,dx = \left|x^2g - xf\right|_{-1}^{1} - \int_{-1}^{1} xg\,dx + \int_{-1}^{1} f\,dx,$$

while an integration over the first equation leads to

$$0 = \int_{-1}^{1} \frac{\partial}{\partial x}(xf-g)\,dx = \left|xf-g\right|_{-1}^{1} = \left|xf-x^2g\right|_{-1}^{1}.$$

Thus with (3·32) we obtain

$$\int_{-1}^{1} xg\,dx = \tfrac{1}{2}, \quad \text{if} \quad \gamma = 0,\ \beta = 0,$$

and hence

$$\tau\frac{dJ}{dt} = -J - \frac{\gamma\beta}{2} + \beta + \gamma\int_{-1}^{1} xf\,dx. \tag{3·35}$$

To eliminate $\int xf$ we notice that, according to (3·30) with (3·34),

$$\tau\int_{-1}^{1} x\frac{\partial f}{\partial t}dx = \int_{-1}^{1} x\frac{\partial}{\partial x}(xf-g)\,dx = -\int_{-1}^{1} xf\,dx + J$$

if we make use of the fact that according to (3·33) $\left|x^2f-xg\right|_{-1}^{1} = 0$. Let us now differentiate (3·35) with respect to t and introduce

$$\frac{d}{dt}\int_{-1}^{1} xf\,dx = \int_{-1}^{1} x\frac{\partial f}{\partial t}dx$$

from the above equation. Using once more (3·35) we can eliminate $\int_{-1}^{1} xf$ and obtain

$$\tau^2\frac{d^2J}{dt^2} + 2\tau\frac{dJ}{dt} + (1-\gamma)\,J = \left(1-\frac{\gamma}{2}\right)\left(\beta + \tau\frac{d\beta}{dt}\right).$$

Finally, in order to obtain P (see (3·9) and (3·34)) we multiply this equation by $\mu n\cos\theta$ and replace $\cos^2\theta$ by 1/3 (polycrystalline material). Introducing β from (3·21) we obtain

$$\tau^2\frac{d^2P}{dt^2} + 2\tau\frac{dP}{dt} + (1-\gamma)\,P = \frac{\mu^2n}{3kT}\left(1-\frac{\gamma}{2}\right)\left(G + \tau\frac{dG}{dt}\right). \tag{3·36}$$

Thus for static fields, $G = G_0$, the equilibrium value P_0 of P becomes, if we neglect γ^2 and higher powers in γ,

$$P_0 = \frac{\mu^2nG_0}{3kT}\left(1+\frac{\gamma}{2}\right). \tag{3·37}$$

For time-dependent fields suppose

$$G = G_0\phi(t), \tag{3·38}$$

where G_0 is time independent and $\phi(t)$ is a function of time. Introducing this and (3·37) into (3·36) we find

$$\tau^2\frac{d^2P}{dt^2} + 2\tau\frac{dP}{dt} + (1-\gamma)\,P = (1-\gamma)\,P_0\left(\phi + \tau\frac{d\phi}{dt}\right). \tag{3·39}$$

Equations (3·37) and (3·39) clearly demonstrate the deviations of the first-order values of the polarization from the zero order (i.e. Onsager's theory), equations

(3·17) and (3·15). We see that the effect of the reaction field is twofold. First, the static polarization is larger by a factor $1+\gamma/2<3/2$ and thus lies between the values according to Onsager and according to Debye. Secondly, the time dependence is different from both Onsager's and Debye's theories. The dipoles are no longer independent of each other and thus reach an equilibrium value no longer $\sim e^{-t/\tau}$ but in a more complicated way. Let us calculate the time dependence of P in a periodic field

$$\phi(t) = e^{i\omega t}. \tag{3·40}$$

Inserting this into (3·39) we find in first order in γ

$$P = \frac{P_0 e^{i\omega t}}{1+i\omega\tau}\left[1-\gamma+\frac{\gamma}{(1+i\omega\tau)^2}\right], \tag{3·41}$$

which for $\gamma = 0$, of course, is identical with the well-known Debye formula.

7. *Liquids with low viscosity.* We have seen that in solids a dipole has two equilibrium positions with opposite dipole direction. An external field alters the transition probabilities between the two positions in such a way that a dipole will prefer one of the two directions thus giving rise to a polarization In liquids, on the other hand, a dipole may have any direction. An external field tends to turn a dipole into a direction parallel to its own, against frictional forces and thermal motion. In spite of this difference from the solid case, the polarization in liquids is in our zero order determined by equations of the same type as in solids ((3·15), (3·17)) as can be seen from the work of Debye (1929). The whole difference lies in the different expressions for the time of relaxation which in solids is connected with the transition probabilities between the two equilibrium positions (see equation (2·1)) while in liquids with low viscosity it can be derived from the viscosity (see § 3).

Taking into account the reaction field, however, low-viscosity liquids behave differently from solids and high-viscosity liquids. We do not want to give details of the calculations because they only show that in low-viscosity liquids the influence of the reaction field is negligible. For the static polarization, for instance, the first-order value of the polarization is found to be $(1+\gamma/24)$ of its zero order value in contrast to a factor $(1+\gamma/2)$ in solids (3·37). Thus it follows that for low-viscosity liquids, Onsager's theory (i.e our zero approximation) is a very good approximation, as is also suggested by the work of Van Vleck (1937*b*). This difference between such liquids and solids is due to the fact that in these liquids a dipole changes its direction more gradually than in solids, thus giving the reaction field a greater chance to adapt itself to the dipolar direction.

IV. Results and discussion

From the calculations of the previous sections we can now derive our final results for the dielectric constant and the dielectric losses. We shall restrict ourselves to a discussion of solids and liquids with high viscosity because, as we saw in § 7, Onsager's theory is a sufficiently good approximation for liquids with low viscosity.

8. *The static dielectric constant* From expression (3·37) for the static polarization we obtain the static dielectric constant ϵ_s, using equations (2·2), (2·3) and (3·3),

$$\epsilon_s = \epsilon_\infty + \frac{\mu^2}{a^3 k T}\left(1+\frac{\gamma}{2}\right)\frac{G_0}{E_0}, \tag{4·1}$$

where E_0 is the static electric field G_0 as well as γ depend on the shape of the cavity For the simplified case of a spherical cavity we can insert G_0 from (3·4) and find

$$\epsilon_s = \epsilon_\infty + \frac{3\epsilon_s}{2\epsilon_s+\epsilon_\infty}\frac{\mu^2}{a^3 k T}\left(1+\frac{\gamma}{2}\right), \tag{4·2}$$

where, according to (3·11) and (3·7),

$$\gamma = \frac{\mu R_1}{kT} = \frac{\mu^2}{a^3 k T}\frac{2(\epsilon_s-\epsilon_\infty)}{\epsilon_\infty(2\epsilon_s+\epsilon_\infty)} < 1. \tag{4·3}$$

Here we can calculate μ^2/a^3kT from equation (4·2) and insert it into (4·3) As usual we have to neglect γ^2 terms, and obtain

$$\gamma = \tfrac{2}{3}\frac{(\epsilon_s-\epsilon_\infty)^2}{\epsilon_s\epsilon_\infty} < 1. \tag{4·4}$$

Thus finally

$$\epsilon_s = \epsilon_\infty + \frac{3\epsilon_s}{2\epsilon_s+\epsilon_\infty}\frac{\mu^2}{a^3 k T}\left(1+\tfrac{1}{3}\frac{(\epsilon_s-\epsilon_\infty)^2}{\epsilon_s\epsilon_\infty}\right). \tag{4·5}$$

It should be remembered that this formula holds for temperatures T larger than a critical temperature T_0, which has to be determined from the condition $\gamma < 1$, which, using (4·4), is equivalent to

$$\epsilon_s < \frac{7+\sqrt{33}}{4}\epsilon_\infty \sim 3\epsilon_\infty. \tag{4·6}$$

This condition thus means that coming from high temperatures where ϵ_s approaches ϵ_∞ equation (4·5) is valid until we reach a temperature T_0, where $\epsilon_s \sim 3\epsilon_\infty$. Nothing can be said from the present theory about the behaviour of ϵ_s for $T < T_0$. Our expression (4·5) for ϵ_s contains two parameters, the high-frequency dielectric constant ϵ_∞ and μ^2/a^3k which has the dimension of a temperature. To compare (4·5) with experimental results we require measurements over a fairly large temperature range which we found in a paper by Garton (1939). He has measured the dielectric constant of a chemically pure synthetic resin (glycol phthalate) between room temperature and about 200° C. Such resins are amorphous substances which gradually soften as the temperature increases but have no well-defined melting-point. According to § 3 we may expect such substances to behave according to our solid model as is actually suggested by the agreement of our theory with the experiments as shown in figure 3. The experiments are carried out at frequencies of 50 kcyc./sec., 50 cyc./sec. and for static fields. They can be used to deduce the static dielectric constant $\epsilon_s(T)$ since the function $\epsilon_s(T)$ forms the envelope to the set of temperature curves $\epsilon_\nu(T)$ where the suffix ν denotes the frequency of the field.

The experiments suggest $\epsilon_\infty \sim 3$. The remaining unknown constant μ^2/a^3k was determined as 1140° abs. by equating theoretical and experimental values for $T = 200°$ C. Figure 3 shows good agreement between theory and experiment. To show the influence of the γ-terms we notice that the factor $1 + \gamma/2$ (see equation (4·2)) is 1·25 at 200° C, and 1·45 at 80° C. In order to come to a final decision about the validity of equation (4·5), more experiments would be required.

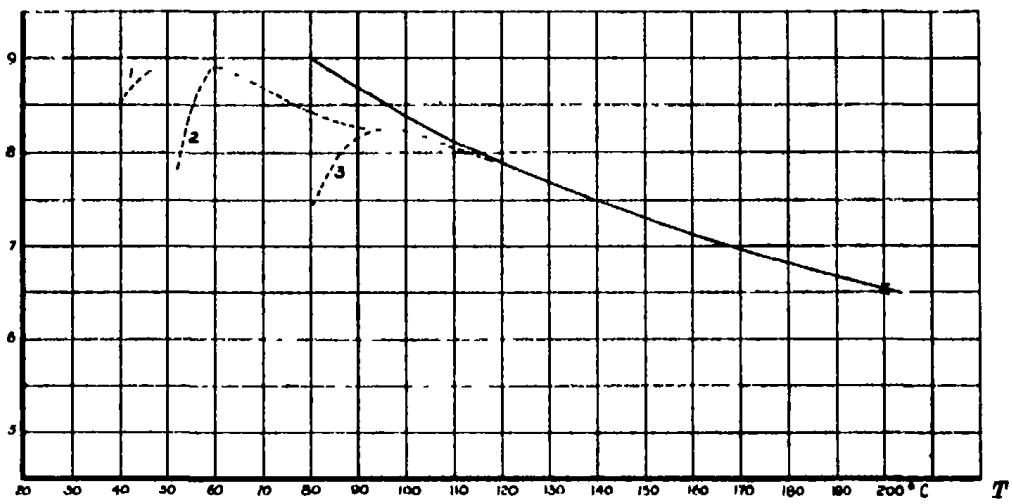


FIGURE 3. Full line. Theoretical curve according to equation (4·5) with $\epsilon_\infty = 3$ as suggested by experiment. The unknown parameter μ^2/a^3k was determined as 1140° K by equating (4·5) with the experimental value at 200° C. Dotted lines: Experimental curves (from Garton 1939), 1, static, 2, 50 cyc./sec.; 3, 50 kcyc./sec. Curve 3 is practically identical with the theoretical curve for $T > 120°$ C. The theory becomes invalid when ϵ reaches $\epsilon = 9$.

9 *Dielectric loss.* In alternating fields we find from (2·2), (2·3), (2·5), (3·41), (3·40), (3·37) and (3·38)

$$\epsilon = \epsilon' - i\epsilon'' = \epsilon_\infty + \frac{\mu^2}{a^3kT}\frac{1}{1+i\omega\tau}\left[1 - \frac{\gamma}{2} + \frac{\gamma}{(1+i\omega\tau)^2}\right]\frac{G}{E}, \tag{4·7}$$

or for spherical cavities, using (3·4),

$$\epsilon' - i\epsilon'' = \epsilon_\infty + \frac{3\epsilon}{2\epsilon+\epsilon_\infty}\frac{\mu^2}{a^3kT}\frac{1}{1+i\omega\tau}\left[1 - \frac{\gamma}{2} + \frac{\gamma}{(1+i\omega\tau)^2}\right], \tag{4·8}$$

where $\gamma < 1$ is now given by (4·4). Separating real and imaginary parts this permits us to calculate ϵ''/ϵ' and hence $\tan\theta$ (see (2·6)). The result would lead to a very complicated formula. Great simplifications can be made, however, if we assume

$$(\epsilon - \epsilon_\infty)\frac{2\epsilon+\epsilon_\infty}{2\epsilon} = \left(\epsilon - \frac{\epsilon_\infty}{2}\right)\left[1 - \frac{\epsilon_\infty^2}{2\epsilon(\epsilon - \epsilon_\infty/2)}\right] \sim \epsilon - \frac{\epsilon_\infty}{2}, \tag{4·9}$$

which is correct unless ϵ approaches ϵ_∞. An elementary calculation shows that with this approximation

$$\tan\theta = \frac{1}{\epsilon'}\left[(\epsilon_s - \epsilon')\left(\epsilon' - \frac{\epsilon_\infty}{2}\right)\right]^{\frac{1}{2}}\left(1 - \frac{\gamma}{2}\right). \tag{4·10}$$

This can be compared with the well-known Debye formula

$$\tan\theta = \frac{1}{\epsilon'}[(\epsilon_s - \epsilon')(\epsilon' - \epsilon_\infty)]^{\frac{1}{2}}. \tag{4·11}$$

The first difference we notice is that for $\gamma = 0$, ϵ_∞ in (4·11) has been replaced by $\epsilon_\infty/2$. This, however, is of minor importance in view of approximation (4·9). In fact we would obtain in (4·10) $\epsilon' - \epsilon_\infty$ instead of $\epsilon' - \epsilon_\infty/2$ if we had assumed $\epsilon - \epsilon_\infty$ to be small. The main difference of our formula (4·10) is thus a reduction of the Debye formula (4·11) by a factor $1 - \gamma/2$ which can take values between 1/2 and 1. In fact, most experimental values of $\tan\theta$ are smaller than required by equation (4·11) (see, for instance, Garton (1939)). Our formula (4·10), therefore, will give better agreement than (4·11). Even so, the theoretical values of $\tan\theta$ are still larger by nearly a factor 2 than those observed by Garton. This probably means that the time of relaxation has at a given temperature a whole range of values as suggested by Garton and others.† This in fact follows at once from our model if we remember that resins are amorphous substances. Different molecules are, therefore, in slightly different positions relative to their neighbours which should give rise to different transition probabilities between the two stable positions of the dipoles We thus require experiments on crystalline solids of the type considered here but unfortunately could not find any publications. We have, therefore, no possibility at present to compare our results on dielectric losses with experiments.

The authors are indebted to the British Electrical and Allied Industries Research Association for permission to publish this paper.

References

Debye, P 1929 *Polare Molekeln.* Leipzig.
Fowler, R. H. 1935 *Proc. Roy. Soc.* A, **149**, 1.
Frohlich, H. 1942 *Proc. Phys. Soc.* **54**, 422.
Fuoss, R. M. & Kirkwood, J. G. 1941 *J. Amer. Chem. Soc.* **63**, 385.
Garton, C. G. 1939 *J. Instn Elect. Engrs*, **85**, 625.
Kurtz, S. S. & Ward, A. L. 1937 *J. Franklin Inst.* **224**, 583, 697.
Muller, A. 1928 *Proc. Roy. Soc.* A, **120**, 437.
Muller, A. 1932 *Proc. Roy. Soc.* A, **138**, 514.
Onsager, L. 1936 *J. Amer. Chem. Soc.* **58**, 1487.
Pauling, L. 1930 *Phys. Rev.* **36**, 430.
Sauer, J. & Temperley, H. N. V. 1940 *Proc. Roy. Soc.* A, **176**, 203.
Ubbelohde, A. R. 1938 *Trans. Faraday Soc.* **34**, 289.
Van Vleck, J. H 1937*a* *J. Chem. Phys* **5**, 320.
Van Vleck, J. H. 1937*b* *J. Chem. Phys* **5**, 556.

† A theoretical analysis of this question has been given by Fuoss & Kirkwood (1941).

Internal stress created by plastic flow in mild steel, and stress-strain curves for the atomic lattice of higher carbon steels

By S. L. Smith, D.Sc., F.C.G.I., *Engineering Department*, and W. A. Wood, D.Sc., *Physics Department*, *National Physical Laboratory, Teddington, Middlesex*

(*Communicated by Sir Charles Darwin, F.R.S.—Received 6 August* 1943)

[Plate 13]

In previous work, stress-strain curves for the atomic lattice of certain metals have been obtained from X-ray diffraction measurements of the lattice dimensions of test specimens under tension or compression, and it has been shown that when the external yield stress is exceeded, there is a systematic departure from Hooke's Law. It is pointed out in the present paper that this departure indicates that the external applied stress above the yield is no longer balanced primarily by simple displacement of the atoms but also by a new type of secondary internal stress brought about by the process of plastic flow, and that this secondary stress, being of a permanent nature, can be measured by the residual lattice strains exhibited by the lattice after removal of the external stress. These residual strains are measured in various directions to the stress direction for mild steel subjected to tension, and it is shown that the lattice after tension exhibits a longitudinal compression and a transverse expansion in the ratio of 2.1, which means that the density of the material is thereby kept constant. Comparisons of X-ray and mechanical measurements further show that the hysteresis loop exhibited by the external stress-strain curve of mild steel after overstrain can disappear and the linear elastic relation be recovered without any corresponding change in the internal stress, which is therefore a more fundamental physical property. It is also shown that when the elastic range is extended by overstrain in tension, there is no symmetrical increase in the elastic range in subsequent compression, thus confirming the existence and direction of the secondary internal stress. Finally, the lattice stress-strain curves are also obtained for a 0.4 % C steel (partially pearlitic) and a 0.8 % C steel (pearlitic), and by comparison with the results on pure iron and 0.1 % C steel (annealed) it is shown that the maximum residual internal strain developed by the lattice increases markedly with the fineness to which the crystallites can be broken down by the plastic deformation.

Behaviour of metallic lattice under stress

In recent work, stress-strain curves have been obtained from X-ray diffraction measurements of the changes in lattice dimensions of certain metals subjected to tension and compression (Smith & Wood 1941, 1942*a*, 1942*b*). The shape of the lattice stress-strain curves has brought out a new feature in the properties of these metals under stress. It is convenient here to refer to this effect in order to explain the purpose of the present experiments and the significance now attached to it in the light of later results. Consider a lattice spacing perpendicular to the applied stress as a specimen is stretched in tension. This spacing should contract as the specimen is extended. It is found that the spacing does contract with applied stress up to the external yield point according to Hooke's Law. In this initial

elastic range, the applied stress is thus balanced by the restoring forces between the displaced atoms. But when the yield stress is exceeded and plastic flow sets in, the following phenomena occur.

(i) The atomic spacing does not continue to contract but remains roughly constant despite continued increase in the applied stress. This means that the excess in applied stress above the yield is no longer balanced by simple displacement of the atoms in the crystalline lattice. The inference is made, therefore, that the process of plastic flow of its own accord has produced a secondary type of internal stress opposing the external stress.

(ii) After removal of an applied stress in excess of the yield, the lattice spacing does not return to its initial value, and the change is in the opposite direction to the change whilst under stress. Thus after the transverse contraction under tension, the same spacing exhibits a residual expansion. The further inference is therefore made that the secondary internal stress produced by the plastic flow is of a permanent nature, and can be measured by its reaction on the crystalline lattice when the applied stress is removed.

(iii) A further feature is the apparent homogeneity of the secondary internal stress It is to be expected that local irregularities of the plastic flow in a polycrystalline metal would result in internal stresses of some heterogeneity. These would vary the lattice spacings from point to point, and affect the X-ray diffraction spectra by causing diffusion of the rings. Such diffusion is in fact observed, having been discussed in detail elsewhere (Wood 1939) But superposed on this is an easily measurable overall change in spacing, which indicates an underlying tendency to homogeneity in the secondary stress system.

Secondary stresses produced by plastic flow in tension

The X-ray measurements in our previous work have referred to a lattice spacing at right angles to the applied stress. The point now investigated is the residual change remaining in a lattice spacing inclined at other directions to an applied tension, including the change in the longitudinal direction.

For these experiments, the mild steel employed in previous work has been used. The lattice spacing measured is the (310), since this can be determined the most accurately by practicable X-ray technique. The measurement of a spacing which will vary from grain to grain according to orientation with respect to the stress direction involves some points which should perhaps be noted. These may be indicated by reference to figure 1, which shows diagrammatically the (310) diffraction cone reflected in a backward direction from the specimen S and meeting the photographic film, which is perpendicular to the incident X-ray beam OS, in the ring with radii OY and OZ in the plane of the paper. The point to note is that when the axis of the specimen is inclined to the incident beam, as drawn, the radii OY and OZ will be unequal, because the (310) planes in the grains reflecting to Y and those reflecting to Z are differently inclined to the axis of the specimen, and thus the

spacings will have been affected differently by the previously applied tensile stress. It will be seen from the diagram that if the specimen is inclined to the beam at an angle α, then measurement of the radius OZ will give the angle of incidence i_2 corresponding to those (310) normals which make the angle $\beta = \alpha - i_2$ with the stress previously applied along the axis. It can be shown also that measurement of the radius (not drawn) perpendicular to the plane of the paper will give the angle of incidence i with those (310) normals for which $\cos\beta = \cos\alpha \cos i$, or since, in the back-reflexion technique here used, i is small, for which $\beta \sim \alpha$. This latter radius and OZ were those actually measured. The corresponding angles of incidence then gave by Bragg's Law a measure of the (310) spacing when inclined at these angles β to the initial stress direction, and measurements for various settings of the angle α gave the change in spacing from the transverse to the longitudinal direction.

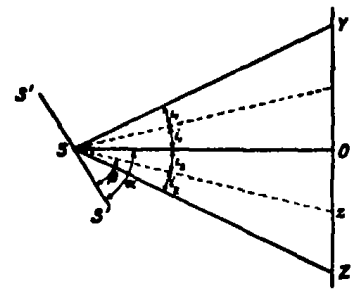


FIGURE 1. $S'SS'$, specimen; OS, incident beam, YSZ, (310) diffraction cone; YOZ, trace of photographic film; Sz, normal to (310) planes reflecting to Z.

In practice, it is easier to determine a diameter than a radius, and OZ was measured as follows. Two exposures were made on one film with the half of the film on the side OY screened; after the first exposure, the film was rotated through 180° in its own plane about O, a movement specially arranged for in the design of the film holder, and a second exposure then given The equatorial diameter of the ring then obtained gave twice the required radius. Alternatively, after the first exposure, the specimen could be rotated into a symmetrical position inclined at angle α on the opposite side of the incident beam, and the second half of the film then exposed The equatorial diameter again gave twice the required radius.

The changes in (310) spacing given in the results below are based, for each setting of α, on the difference between the radius of the diffraction ring from the stressed specimen and the radius of the ring obtained under the same conditions from a specimen before stressing. This comparison method was considered to reduce possible experimental errors to a minimum. By preliminary calibration and calculation, it was known that a change in diameter of 1 mm. in the diffraction ring corresponded to a change in (310) spacing of 0·04 %, which was adequately sensitive for measuring the changes involved. For this, the specimen-film distance was 10 cm., and the incident wave-length was the cobalt $K\alpha_1$.

Results

In figure 2 is plotted the residual change in (310) spacing from the transverse to the longitudinal directions for a tensile specimen which had been stressed to 20 tons/sq. in., the yield point of this material being 16 tons/sq in. Specimens stressed to other values beyond the yield exhibited the same essential features, so that it is necessary only to consider in detail the example given. The closest approach to the longitudinal direction which could be obtained from measurement of the equatorial radius OZ of the diffraction ring was when the (310) spacing was inclined at 11° to this direction. Measurement of the radius perpendicular to OZ permitted an estimate to be made of the (310) spacing when inclined at 5°.

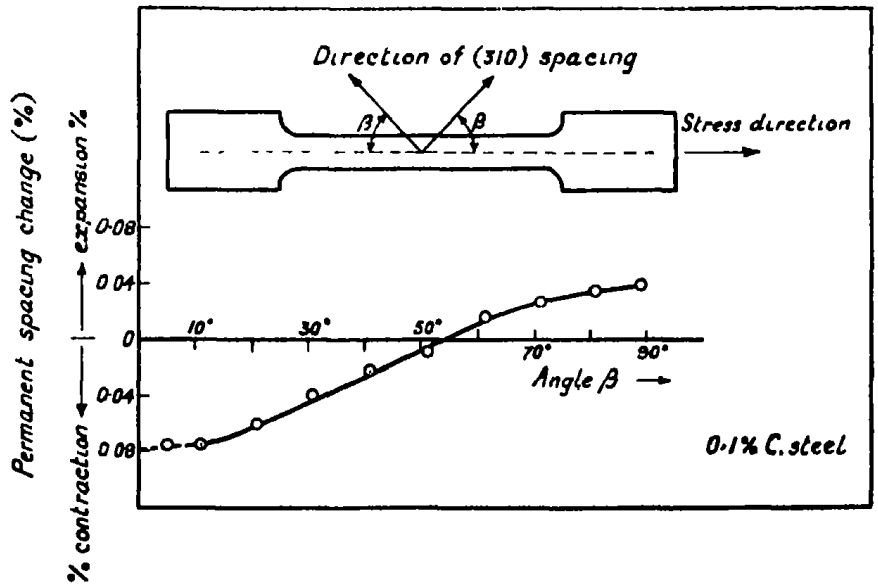


FIGURE 2. Variation in permanent lattice strain (remaining after removal of tensile stress) for different inclinations β of the (310) spacing to the initial stress direction. The residual longitudinal strain ($\beta = 0°$) is found to be twice the residual transverse strain ($\beta = 90°$)

It will be seen from the curve that at right angles to the stress direction the lattice spacing is left with a residual expansion, as previously noted. When the spacing becomes inclined to the stress direction, this residual expansion decreases in value and becomes zero at an inclination between 50 and 60°. The spacing then shows a contraction which increases in a regular manner, reaching a maximum as the original stress direction is approached. From extrapolation, and from consideration of the value of the angle where the spacing changes from expansion to contraction, it was found that the magnitude of the contraction in the longitudinal direction was, within limits of measurement, double the value of the expansion in the transverse direction. This means that the residual changes in spacing remaining after application of the tensile stress are so related as primarily to keep the density of the crystalline material constant. Also, if the residual spacing changes are considered to be due to an equivalent system of secondary stresses created by the

plastic flow in opposition to the external applied stress, then the internal stress produced by the applied tension is not a simple compression in the reverse direction; for the longitudinal and transverse lattice changes would then be related according to Poisson's ratio, which for this material is 0·27. The equivalent secondary stresses must consist of a longitudinal compression together with a transverse tension.

Effect of secondary internal stresses on elastic properties

If, as suggested by the above work, plastic flow produces a system of internal stresses, two questions then present themselves

(*a*) The elastic properties of mild steel after overstrain are known to change with time Are the internal stresses permanent or do they decrease with time as the material recovers its elasticity? (Mechanical hysteresis.)

(*b*) To what extent will the internal stresses produced by tension affect the elastic properties in compression? (Bauschinger effect.)

These points were examined in the following experiments.

(*a*) *Mechanical hysteresis*

A characteristic of a metal stressed beyond the yield is that the external stress-strain relation obtained on unloading and reloading then forms a loop. In low-carbon steels, such as that under examination, the loop usually closes with time into a straight line, when the material may be said to have recovered its elasticity (Smith & Howard 1932). The point tested was whether such recovery was accompanied by changes in the residual lattice strain effect.

The observations, which extended over three months, showed that the hysteresis loop for the mild steel (and also for pure iron) could close up without any accompanying change whatever in the X-ray diffraction patterns. Thus in a typical normalized mild steel stressed to 24 tons/sq. in. (yield point 20 tons/sq. in), the width of the hysteresis loop was halved after 17 hr at room temperature and fully closed after $4\frac{3}{4}$ days, when the material thus became elastic over the range 0–24 tons/sq. in. The residual lattice strains, on the other hand, were unchanged after the three months. Also experiments made on specimens overstrained and then heated to various temperatures showed that recovery from hysteresis was completed after about 1 hr. at temperatures of 100–200° C, whereas recovery from the residual lattice strains required heat treatment at 400–500° C for pure iron and 500–600° C for the mild steel. The results therefore indicate that recovery of elasticity after overstrain is not associated with any regular modifications of crystalline structure, and is probably due to a general smoothing out of local irregularities, produced in the initial plastic flow, which are not sufficiently extensive to affect the X-ray diffraction spectra. The permanent lattice strain is associated with more fundamental modifications of mechanical properties.

(*b*) *Bauschinger effect*

In the annealed mild steel used the initial elastic range was approximately the same in tension as in compression, 16–17 tons/sq. in According to the X-ray results, overstrain beyond the yield in tension creates secondary internal stresses in the direction of compression, and these are in practice permanent at room temperature. The subsequent external elastic range in compression after overstrain in tension should therefore be reduced. It will be recalled that this suggestion was put forward by Bauschinger in the middle of the last century on the basis of extensive mechanical tests.

For a proper examination of the point, it was considered necessary that the hysteresis loop produced in the tensile cycle should be closed to a straight line before the subsequent measurements in compression were undertaken The conditions required for this recovery of elasticity after overstrain in tension, without affecting the permanent lattice strain, were known from the work referred to in the preceding section.

The first tests were made on a cylindrical tensile specimen, of diameter 0·358 in. and parallel test length 1 in This was loaded in tension to 24 tons/sq in , and allowed to become stabilized by standing at room temperature, the material then becoming elastic from 0 to 24 tons/sq. in A $\frac{3}{4}$ in. length was then cut from the parallel portion to give a cylinder suitable for testing in compression

Mechanical measurements. These showed that the limit of proportionality, which was 24 tons/sq in. in tension, became 15 tons/sq in in compression instead of the 24 tons/sq in which might have been expected if the process of strain-hardening were non-directional

X-ray measurements. The residual lattice strain was for convenience measured in the direction at right angles to the applied stress. After the load in tension, this residual transverse strain for the (310) spacing was +0·06 % (expansion). During compression the load was removed at the stresses indicated below, and the residual transverse strain redetermined with the following results:

compression	residual (310) strain
after 4 tons/sq. in.	+0 06 %, still expanded
8 5	+0 06 %, unaltered
12·8	+0 06 %, unaltered
17 0	+0 04 %, decreasing
20·3	+0·02 %, marked decrease
23·5	−0 01 %, contraction
26·1	−0·04 %, marked contraction

From previous measurements on the lattice stress-strain curve in tension, it is known that after the initial loading in tension the lattice stress-strain relation is linear over the range 0–24 tons/sq. in. (vide figure 3*b*). The above observations show that (i) this elastic range of the lattice in tension is reduced in subsequent compression to a range somewhat less than 17 tons/sq. in., which is of the same order as the reduction in the external mechanical elastic range; (ii) when this reduced

elastic range in compression is exceeded, the subsequent plastic flow produces the directional effect of first slowly cancelling out the residual lattice strain left after the previous tensile stress, and then reversing its direction, so that the transverse (310) spacing, which was previously expanded, becomes finally contracted.

Both the mechanical and the X-ray measurements therefore confirm that the plastic flow in tension appears to create a secondary internal stress in the direction of compression.

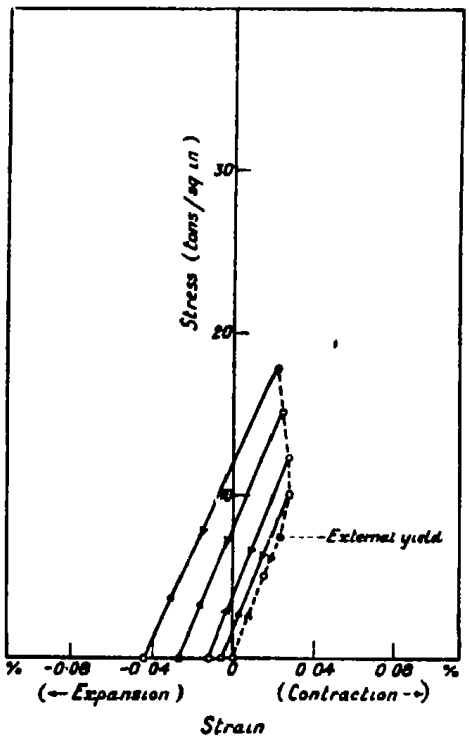


FIGURE 3*a*. Lattice stress-strain curve for iron (purity 99·95 %).

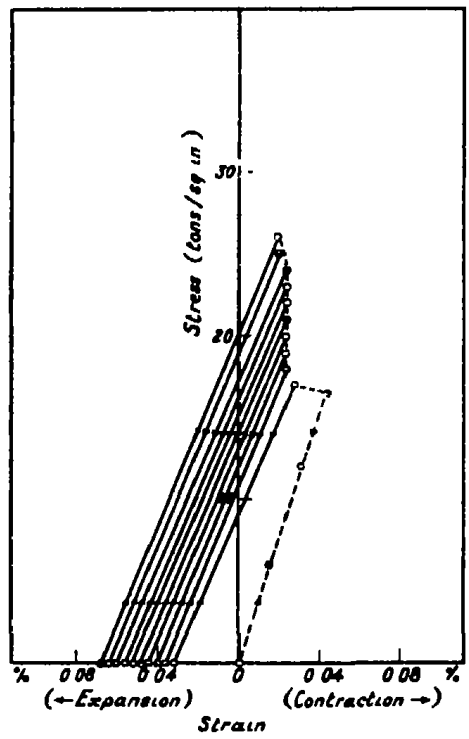


FIGURE 3*b*. Lattice stress-strain curve for 0·1 % C steel.

Figures 4*a*, 4*b* (plate 13) compare the (310) spacing after 24 tons/sq. in. tension with the same spacing after the subsequent 23·5 tons/sq. in. in compression. These are reproduced to show that the difference in diameter of the respective diffraction rings is of an easily measurable order; and the difference clearly demonstrates the non-elastic behaviour of the metallic lattice under conditions producing plastic flow.

Before concluding this section, reference may perhaps be made to two further confirmatory experiments of the same nature as the above but on mild steel in the normalized, as distinct from the annealed, condition. In these experiments short, thick specimens were used, specially designed to be suitable for tension and compression without the necessity of cutting the specimen. The yield point in the normalized condition was 21 tons/sq. in. The first specimen was loaded to 24 tons/sq. in. in tension, and allowed to become stabilized by resting as before. On testing

in compression, the elastic range was found to be 15 tons/sq. in The second specimen was loaded only to the yield point in tension, and held there for 5 min. until yield was complete. The specimen, after stabilization as before, was tested in compression. The elastic range was reduced to 11–12 tons/sq. in. The breakdown of the grains during plastic deformation, which is one factor to be considered in strain-hardening, can hardly be a 'directional' effect; it would have the same influence in tension or compression. But that the internal stress is directional appears to follow definitely

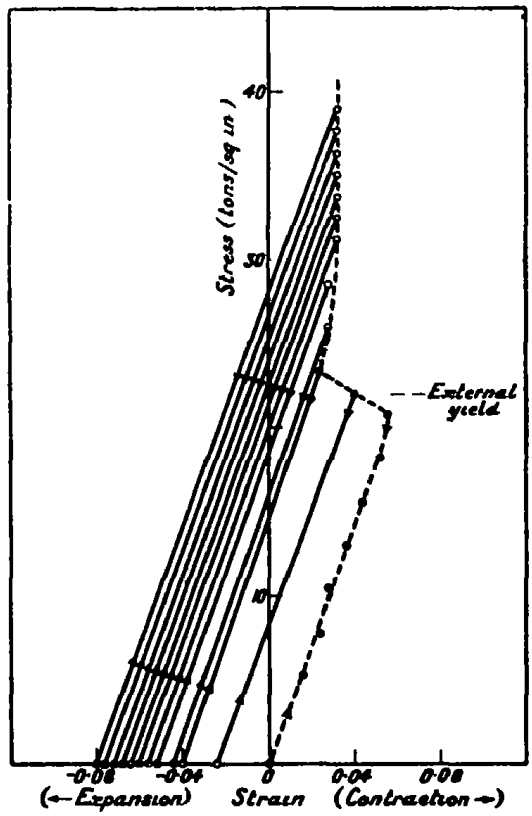


FIGURE 3c. Lattice stress-strain curve for 0·4 % C steel

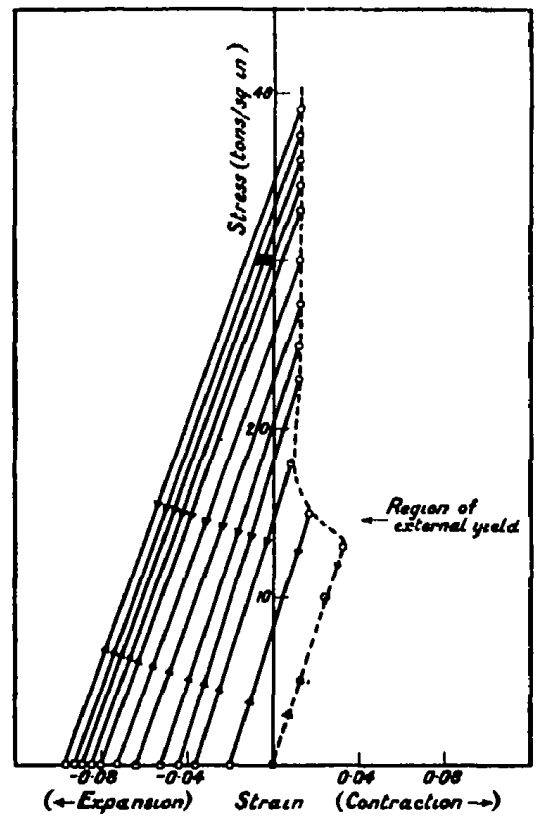


FIGURE 3d. Lattice stress-strain curve of eutectoid steel.

from both the X-ray and the mechanical tests. The increase in elastic range which can be produced by overstrain in tension applies only to the tension cycle and not to the subsequent compression, both externally and for the lattice.

LATTICE STRESS-STRAIN CURVES FOR HIGHER CARBON STEELS

The next step was to consider experiments on the source of the secondary internal stresses.

A possibility is that they arise in a distorted envelope formed around the fragmented grains and crystallites produced by plastic flow. If so, they should be

affected by the texture of the iron or ferrite matrix, and the fineness to which this texture can be reduced by mechanical deformation.

It is not possible to produce crystallites less than a certain size by plastic deformation of a pure metal (Wood 1939). But the texture can be varied by introduction of foreign atoms. The dispersion of carbides in steels, for example, permits of particularly wide variation in the condition of the ferrite matrix Advantage was taken of this fact in comparing the lattice stress-strain relationship for the following materials.

(i) Pure iron, annealed, initial grain size 10^{-2}–10^{-3} cm.

(ii) 0·1 % C steel, annealed, ferrite grains 10^{-2}–10^{-3} cm. and carbon in spheroidized condition.

(iii) 0·4 % C steel, mixture of ferrite grains as above and laminar pearlite.

(iv) 0·8 % C steel, entirely laminar pearlite of bands approximately 0·5 mm at 500 magnification

The above range was purposely chosen also in view of other X-ray work which showed that the maximum diffusion of the diffraction rings reached during deformation varies considerably for steels of different complexity Thus, in pure iron, the sensitive (310) $\alpha_1\alpha_2$ doublet obtained with cobalt K radiation is still strong and well resolved after the severest practicable deformation The same doublet, however, from martensitic, or quenched and tempered steels, is not only unresolved but so diffuse as to be hardly visible. The above range was known to produce an increasing scale of diffusion in the order given, as a result of deformation; but not so extensively as to preclude estimation of the overall changes in diameter of the diffraction rings.

Results

The results refer to the change in (310) spacing transversely to the stress direction, and were obtained in the combined X-ray spectrometer and tensile testing machine already described in previous papers. The curves for the pure iron and 0·1 % C steel were similar to those obtained in the previous work, but, for convenience of comparison, they are reproduced herewith on the same scale as the corresponding curves for the 0·4 and 0·8 % C steels, the set being given in figures 3*a* to 3*d*.

The curves show (*a*) a similar linear stress-strain relation up to the external yield, (*b*) a similar characteristic departure from Hooke's Law as the yield is exceeded, and (*c*) the building up of permanent lattice strains above the yield which remain when the applied stress is removed.

The point of interest in connexion with the present aims is the difference in magnitude of the residual lattice strain developed at the ultimate stress for the different materials. This increases with the complexity of the steels, being 0·04 % for the iron, 0·06 % for the 0·1 % C steel, 0·08 % for the 0·4 % C steel, and 0·09 % for the 0·8 % C steel As already mentioned, the diffusion of the diffraction rings also increases in the same order on deformation of these materials. Now it has been shown recently, by a new X-ray diffraction result (Wood 1943), that this

diffusion in the case of the steels is due almost entirely to reduction of the ferrite grains to fine crystallites, when, with standard X-ray techniques, the number of atomic planes in each crystallite becomes too small to produce sharp reflexions. The crystallite size decreases as the diffusion increases. These steels therefore represent a range in which the ultimate crystallite size is progressively smaller from the iron to the 0·8 % C steel; measurements based on the usual formulae connecting diffusion and particle size gave $3{\cdot}2 \times 10^{-5}$ cm (Fe), 2×10^{-5} cm (0·1 % C steel), $1{\cdot}5 \times 10^{-5}$ cm. (0·4 % C steel) and 10^{-5} cm. (0·8 % C steel) approximately. The maximum residual lattice strain therefore appears to increase as the ultimate crystallite size for the iron matrix is decreased. This result thus suggests a relation between the secondary internal stresses and the surface boundary conditions which would become more effective with reduction in crystallite size.

Discussion

A specimen which is not homogeneous, or which is subjected to a stress distribution that is not homogeneous, could be stressed to a stage where some regions would be plastically deformed while the remainder was within the elastic range of the material. When the external stress is released, the specimen would be in a state of internal stress because of the restraint of the permanently deformed regions on the remainder. These may be termed 'layer' stresses, because they are commonly met with in specimens where the surface has been deformed more than the interior.

It is important to note that such large-scale layer stresses could not explain the internal stresses investigated in the present work. For, in the first place, there is no reason to suppose that the stress distribution across a cylindrical tensile specimen would be so far from homogeneous as to produce effects of the order observed. Also, even if the distribution were not homogeneous, the lattice spacing at the surface would change in a unidirectional manner as the total load was continually increased, whereas in actual fact this is not so, the transverse spacing of mild steel, for instance, exhibits a sudden contraction at the external yield. The point can be settled, however, by etching a specimen and determining the residual lattice strain from the surface to the interior, a procedure possible because of the low penetration of the wave-lengths used in the X-ray diffraction work. This test was undertaken specially on a cylindrical tensile test piece of mild steel 9·75 mm. diameter. After stressing beyond the yield to 24 tons/sq. in., measurements were made of the residual transverse strain shown by the (310) spacing at the surface and at stages as the original surface was etched away in dilute nitric acid. No appreciable change could be observed. A further etching test was made on a similar tensile test piece which was sectioned so that the longitudinal lattice strain also could be measured. This process involves some difficulties because the action of sawing appears to leave an expanded lattice in the surface layers, but it was found that these could be removed by careful turning and etching. It was then observed that after continued deep etching, the longitudinal lattice spacing was contracted

and the transverse spacing expanded always in the same ratio. The difference in corresponding diameters of the diffraction rings is illustrated in figures 5 *a*, *b* (plate 13). These diffraction rings also illustrate the characteristic diffusion produced by the plastic deformation.

The internal stresses here observed are therefore not due to large-scale layer stresses. Analogy with layer stresses suggests, however, a means by which they might arise, though in a more fundamental manner. Thus, it is reasonable to suppose that during plastic flow a crystallite becomes bounded by atomic layers in which the regular crystalline formation becomes highly distorted, or even lost. If the elastic range of this boundary material is greater than that of the crystalline interior, then as the yield is exceeded the boundary may still be deformed elastically whilst the interior undergoes permanent deformation by the plastic flow. On removal of the applied stress, the boundary layer will be unable to recover entirely, because of the permanent deformation of the interior. It will therefore exert a stress with both longitudinal and transverse components on the enclosed crystalline lattice. This would produce residual lattice strains of the type observed. The example just given would be an extreme case. Actually it would be necessary only to suppose that the stress-strain relation for the boundary layers was different from that of the crystalline interior in order to produce the residual strains in the crystalline lattice. We therefore suggest the possibility of a fine-scale distribution of layer stresses, originating in this manner, being an inherent consequence of plastic deformation in a metal.

Finally, the extension of the lattice measurements from the transverse to the longitudinal direction confirms the general principle previously suggested: *after the yield point, the lattice systematically assumes a permanent strain in such a sense as to oppose the elastic strain induced by the applied stress, and of a magnitude which increases in a regular manner with the external stress.*

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

References

Smith, S. L. & Howard, J. V. 1932 *Proc. Roy. Soc.* A, **137**, 519.
Smith, S. L. & Wood, W. A. 1941 *Proc. Roy. Soc.* A, **178**, 93.
Smith, S. L. & Wood, W. A. 1942*a* *Proc. Roy. Soc.* A, **179**, 450.
Smith, S. L. & Wood, W. A. 1942*b* *Proc. Roy. Soc.* A, **181**, 72.
Wood, W. A. 1939 *Proc. Roy. Soc.* A, **172**, 231.
Wood, W. A. 1943 *Nature, Lond.*, **151**, 585.

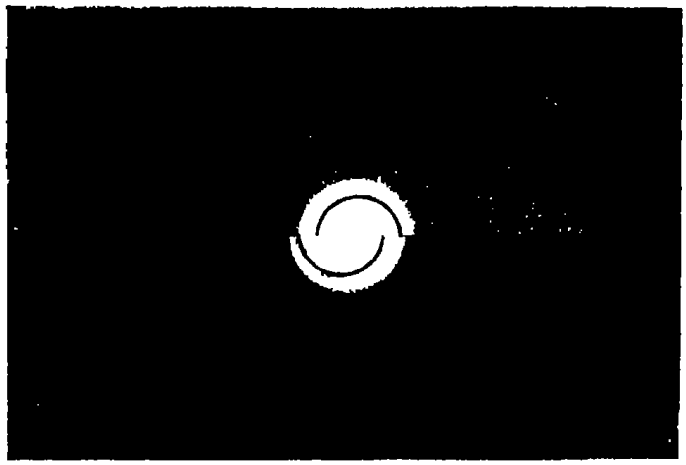


FIGURE 1 Showing non-elastic behaviour of lattice in tension-compression cycle.



FIGURE 5 Showing difference in diameter of (310) rings in longitudinal and transverse directions after plastic flow under tension.

(*Facing p.* 114)

The structure of electro-deposited chromium. A correction.

By W. Hume-Rothery, F.R.S. and M. R. J. Wyllie

(*Received* 14 *January* 1944)

In our recent paper (Hume-Rothery & Wyllie 1943) on the structure of electro-deposited chromium, figures were given for the stress in electro-deposits prepared on steel strip. These figures were deduced by means of an equation given by Stoney (1909) for the relation between the radius of curvature of the steel strip and the stress in the deposit. According to Stoney the radius of curvature, R, of a strip of length l is related to the deflexion of the strip, z, by the equation $R = l^2/8z$. In our work, we assumed that if the strip were held at one end, during electro-deposition, the deflexion was the movement of the free end. Actually, as can be seen from figure 1, if z is the movement of the free end of the rod, then for small deflexions $R = l^2/2z$. If the deflexion is defined as the perpendicular distance from the mid-point of the rod to the chord of the circle joining its two ends, i.e. the distance y in figure 1, then the relation is $R = l^2/8y$, which is that given by Stoney.

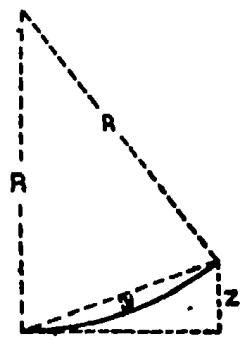


Figure 1.

The stresses given in figure 5 of our paper are thus all four times greater than they should be, and although the general conclusions are quite unaltered, the maximum stress is really only 27·5 tons/in.^{2}, and not 110 tons/in.^{2} as was erroneously stated. This mistake was pointed out to us by Mr G. R. Gardam of the Research Department, Woolwich, and we can only express our apologies for the confusion which may have been caused.

References

Hume-Rothery, W. & Wyllie, M. R. J. 1943 *Proc. Roy. Soc.* A, **181**, 331.
Stoney, G. G. 1909 *Proc. Roy. Soc.* A, **82**, 172.

The three infinite harmonic series and their sums (with topical reference to the Newton and Leibniz series for π)

By F. Soddy, F.R.S.

The following corrections refer to the above paper which was printed in part 989 of this volume

p. 114, third line from end of Abstract *for* 'If $d/2$ is even, . .' *read* 'If $d/2$ is odd, .', and also on p. 124, at the end of the second line preceding formula (17).

p 119, 19 lines from end, d/a should be a/d. The first part of the next sentence should be: 'Integral changes of a/d, in (5) and (6), change the denominators of the terms from $d(M+a/d)$ to, say, $d(M+n+a/d)$, that is, .' In the rest of the paragraph, for (3) read (5), and for (4) read (6).

p 122, l 10, the numerator should be $(-1)^{M+1}$, and, in l. 11, the denominator $(Md)^{2m+1}$ should be $(Md)^{2m-1}$.

INDEX TO VOLUME 182 (A)

Adsorption heat of long-chain compounds and their effect on boundary lubrication (Frewing), 270
Alloy of copper, iron and nickel, dissociation (Daniel & Lipson), 378.
American scientists, organization for the war (Compton), 1.
Anniversary address (Dale), 217.
Asymptotic formulae relating to the physical theory of crystals (Ledermann), 362

Bacterium lactis aerogenes, induced changes in morphology (Hinshelwood & Lodge), 319.
Bauer, S. G. Efflux through a circular orifice, 347

Carbon monoxide flame spectrum III. The cool flame (Gaydon), 199.
Chang, T. S The impulse-energy tensor of material particles, 302.
Chromium structure (Hume-Rothery & Wyllie), 415.
Compton, K. T *Pilgrim Trust Lecture Organization of American scientists for the war*, 1.
Coumoulos, G. D The electron diffraction by amorphous polymers, 166
Crystals, asymptotic formulae relating to the physical theory (Ledermann), 362.

Dale, Sir Henry *Anniversary address*, 217.
Daniel, V. & Lipson, H The dissociation of an alloy of copper, iron and nickel Further X-ray work, 378.
Darwin, Sir Charles The refractive index of an ionized medium, II, 152
Dielectric properties of dipolar substances (Frohlich & Sack), 388.
Diffusion equation in a turbulent medium (Sutton), 48.

Efflux through a circular orifice (Bauer), 347
Eisenschitz, R. Matrix theory of correlation in a lattice, I & II, 244, 260.
Electron diffraction by amorphous polymers (Coumoulos), 166.
Electrons, scattering and polarization by heavy elements (Mohr), 189.
Evaporation from a plane, free-liquid surface into a turbulent air stream (Pasquill), 75.

Fibrous proteins, combination with acids (Gilbert & Rideal), 335.
Frewing, J J. The heat of adsorption of long-chain compounds and their effect on boundary lubrication, 270
Frohlich, H & Sack, R Dielectric properties of dipolar substances, 388.

Gaydon, A. G. The band spectrum of nitrogen. new singlet systems, 286.
Gaydon, A. G. The flame spectrum of carbon monoxide, III. The cool flame, 199.
Gilbert, G. A. & Rideal, E. K. The combination of fibrous proteins with acids, 335
Gimpel, I. & Richardson, Sir Owen. The secondary electron emission from metals in the low primary energy region, 17.

Harmonic series, sum (Soddy), 113, 416.
Hinshelwood, C. N. & Lodge, R. M. A physico-chemical study of some induced changes in the morphology of *Bacterium lactis aerogenes*. A theory of the balance and adaptive variation of certain enzyme processes in bacteria (Abstract), 319.

Howell, H. G. The ultra-violet spectra and electron configuration of HgF and related halide molecules, 95.
Hume-Rothery, W. & Wyllie, M. R. J. The structure of electro-deposited chromium (a correction), 415.
Huxley, H. & Lowery, H. The refraction and dispersion of sulphuretted hydrogen, sulphur dioxide and carbon oxysulphide, 207.

Impulse-energy tensor of material particles (Chang), 302.
Internal stress created by plastic flow in mild steel, and stress-strain curves for the atomic lattice of higher carbon steels (Smith & Wood), 404.

Jánossy, L. & Rochester, G. D. The production of penetrating showers, 180.

Ledermann, W Asymptotic formulae relating to the physical theory of crystals, 362
Lipson, H. *See* Daniel & Lipson.
Lodge, R. M. *See* Hinshelwood & Lodge.
Lowery, H. *See* Huxley & Lowery.

Matrix theory of correlation in a lattice (Eisenschitz), 244, 260.
Mohr, C. B. O. The scattering and polarization of fast electrons by heavy elements, 189.

Nitrogen band spectrum: new singlet systems (Gaydon), 286.

Pasquill, F. Evaporation from a plane, free-liquid surface into a turbulent air stream, 75.
Pebbles, natural and artificial. Their shape under various conditions of abrasion (Rayleigh), 321.
Penetrating showers production (Jánossy & Rochester), 180.
Pilgrim Trust Lecture. Organization of American scientists for the war (Compton), 1.
Presidential address (Dale), 217.

Rayleigh, Lord. Pebbles, natural and artificial. Their shape under various conditions of abrasion, 321.
Refractive index of an ionized medium (Darwin), 152
Relaxation methods applied to engineering problems (Southwell & Vaisey), 129.
Richardson, Sir Owen. *See* Gimpel & Richardson.
Rideal, E. K. *See* Gilbert & Rideal.
Rochester, G. D. *See* Jánossy & Rochester.

Sack, R. *See* Fröhlich & Sack.
Secondary electron emission from metals in the low primary energy region (Gimpel & Richardson), 17.
Smith, S. L. & Wood, W. A. Internal stress created by plastic flow in mild steel, and stress-strain curves for the atomic lattice of higher carbon steels, 404.
Soddy, F. The three infinite harmonic series and their sums (with topical reference to the Newton and Leibniz series for π), 113, 416.
Southwell, R. V. & Vaisey, G. Relaxation methods applied to engineering problems. VIII. Plane-potential problems involving specified normal gradients, 129.

Sulphuretted hydrogen, sulphur dioxide and carbon oxysulphide, refraction and dispersion (Huxley & Lowery), 207.
Sutton, W. G. L. On the equation of diffusion in a turbulent medium, 48.

Ultra-violet spectra and electron configuration of HgF and related halide molecules (Howell), 95.

Vaisey, G. *See* Southwell & Vaisey.

Wood, W. A. *See* Smith & Wood.
Wyllie, M. R. J. *See* Hume-Rothery & Wyllie.

I A R I 75

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY
NEW DELHI

Date of issue	Date of issue	Date of issue

S C P—1/8/47 P J 3 5 48 2000